Voluntary Sustainability Standards and Farm Revenue:

The Pathways to Success?

Abstract

Whether Voluntary Sustainability Standards (VSS) - such as Fairtrade, GlobalGAP or Organic - may contribute to economic sustainability has been subject to vigorous debate over the past decades. To analyze the welfare effects of VSS for family farms, previous studies rely on single equation models which enables analyzing *if* (some) VSS affect welfare, but not *how* (some) VSS affect welfare – or not. Hence, little is known about the mechanisms and the relative importance of these different mechanisms by which VSS contribute to farmer welfare. In this paper, we analyze the channel of effects through which VSS affect net revenue of family farms in Peru. We use nationwide data from Peru's National Agricultural Survey covering five years, multiple agricultural commodities, and multiple VSS. We apply a multiple mediation model which allows to disentangle the main revenue-determining mechanisms and to compare their relative importance. We find that prices (and not yields) are the main channel through which VSS affect farmer welfare. However, higher prices cannot offset higher production costs, resulting in zero net welfare gains. We do not find an effect through yields but identify a large potential effect on net revenue. We find heterogeneity in the effects by standard and crop, with crops certified to standards that apply a system of quality-based price differentiation having the largest impact on farmer welfare through a price effect. From a policy perspective, we highlight potential improvements in standard VSS design to effectively improve economic sustainability.

Keywords: private standard, certification, pathway, simultaneous equation model, mediation, indirect effect

1 Introduction

A major challenge for current and future policy is to keep economic activities within planetary boundaries while meeting societal needs (Fanning et al., 2021). This is particularly challenging in the agrifood sector as the sector is a main contributor to greenhouse gas emissions, climate change, deforestation, biodiversity loss, resource depletion and pollution (Chrysafi et al., 2022). At the same time, more than two-thirds of people living in extreme poverty are farmers or farm workers, and global hunger has been on the rise since 2015 after decades of progress towards the zero-hunger goal (FAO et al., 2020; Pharo et al., 2019). A variety of strategies have been proposed to improve sustainability in the global food system, ranging from government regulations to voluntary market-based instruments. The latter includes private or voluntary sustainability standards (VSS), such as Fairtrade, GlobalGAP or Organic, which intend to increase efficiency (e.g., by reducing information asymmetries or negative externalities) and equity (e.g., by redistributing rents) in global value chains (Swinnen et al., 2021). VSS adoption by producers, processors or traders is voluntary, and compliance is usually controlled and enforced through third-party certification.

VSS are of significant international food policy relevance, considering their substantial growth over the past decade in terms of certified farmers, acreage, and products. They now cover at least 8% of the global cropland (Willer et al., 2022). At the same time, consumer willingness to pay for certified food is growing; food and retail companies are increasingly relying on VSS in their corporate sustainability and product differentiation strategies; and national governments, donors and civil society organizations are increasingly promoting VSS projects and financing smallholder certification schemes (UNFSS, 2022). The increasing policy importance of VSS is documented by the recent emergence of VSS in trade agreements, public procurement policies, due diligence requirements, market-access regulations, export promotion policies, and international initiatives such as the UN Forum on Sustainability Standards (UNCTAD, 2023).

Understanding what VSS mean economically for farmers is pertinent, particularly for low- and middle-income countries, where VSS are most commonly adopted (Tayleur et al., 2017). Despite the growing body of evidence, substantial gaps remain in our understanding of the welfare impact of VSS adoption in tropical commodity sectors, which hinders the VSS policy dialogue. Review studies show that, depending on the context, VSS provide no or only a small economic benefit to producers (DeFries et al., 2017; Meemken, 2020; Oya et al., 2018; Traldi, 2021), but surprisingly little research has been done on *how* VSS adoption affects producer welfare. Some argue that prices are (or should be) the main mechanism through which VSS (can) affect farmer welfare (Beuchelt and Zeller, 2011; Estrella et al., 2022; Mitiku et al., 2017; Qiao et al., 2016), while others hint at higher yields and sales volumes (Akoyi and Maertens, 2018; Barham et al., 2011; Barham and Weber, 2012; Ruben and Fort, 2012; Vanderhaegen et al., 2018). Most VSS impact studies are case-specific which limits the internal (e.g., through the ability to address selection bias) and external validity (i.e., the ability to generalize). Understanding the mechanisms through which VSS affect farmer welfare is important because VSS can only contribute to sustainable agricultural development if their adoption is economically beneficial for producers.

This study aims at understanding how VSS affect farmer welfare by (1) exploring the main revenue-determining mechanisms through which VSS affect net farm revenue among small- and medium-scale farms in Peru and disentangle their relative importance, (2) investigating heterogenous revenue effects by VSS and (3) by crop. We rely on nationally representative farm-household survey data from Peru over the period 2015-2019 and multiple mediation models. We focus on small- and medium-scale farms (hereafter referred to as "family farms")[i], representing the majority (98%) of Peruvian farms (INEI and DNCE, 2023)

We make an innovative contribution to the VSS impact literature by looking beyond the direct

[i] We follow the Peruvian government's definition, which defines small and medium-sized farms as family farms of generally less than 50 hectares.

association between VSS and economic welfare. We specifically focus on the mechanisms through which the effect occurs and the relative importance of these mechanisms using mediation models. Previous quantitative impact studies that investigate multiple outcome indicators rely on single equation models, which allows for analyzing different impact mechanisms but not the relative importance of these mechanisms or possible counterbalancing effects. In addition, we use a nationally representative dataset from Peru that covers a five-year period (2015-2019), multiple agricultural commodities, and multiple VSS, which allows us to improve upon the internal and external validity of estimates in most VSS impact studies. With the exception of Meemken (2021) using the same dataset, previous studies mostly use cross-sectional data and focus on a specific crop, area and/or VSS, which complicates understanding the complexity of VSS, examining dynamics over time, addressing selection bias, and generalizing conclusions (e.g. Estrella et al., 2022; Qiao et al., 2016; Vanderhaegen et al., 2018).

Moreover, we make a methodological contribution by estimating mediation models using longitudinal data, which only few economic studies have done so far (Fakhfakh and Jarboui, 2020; Wahba and Elsayed, 2015). While the advantage of using longitudinal data in structural equation modeling is the use of more flexible panel data estimators (Bollen and Brand, 2010), very few papers have estimated mediation models using panel data methods other than pooled OLS, thereby relying on a strong exogeneity assumption that is often hard to meet in applied research due to selection bias.

Finally, the focus on Peru, where agriculture is important in GDP (11%, of which 67% is due to agricultural exports) and employment (27.5%), and where 40% of family farmers live below the international poverty line of $1.90 per day (FAO, 2023; INEI and DNCE, 2023; The World Bank, 2023), is particularly relevant. Over the past decade, Peru's certified area has increased enormously, for example by 128% per year for cocoa, 33% for coffee and 36% for bananas (ITC, 2023). Currently, 11% of producers are certified, making the country the 12th largest certified country by number of producers. Moreover, the country's size and high agrodiversity ensure the presence of a wide variety

of crops and VSS, allowing the analysis of impacts on multiple crops and VSS (ITC, 2023). In addition, the current VSS literature is biased towards studies focussing on Africa (Traldi, 2021).

2 Literature review

2.1 Conceptual framework

VSS can be seen as a bundle of interventions and associated organizational changes, resulting, directly or indirectly, in different intermediate outcomes and long-term sustainability impacts. To effectively monitor their effects, VSS articulate theories of change (ToCs), which are structured process diagrams linking interventions to impact. We review the ToCs for 14 leading VSS according to Willer and co-authors (2022), to create insights and hypotheses about the impact pathways through which VSS can affect farm welfare, via its impact on both certified and non-certified produce (Figure 1). It should be noted that some pathways may interact because multiple interventions are taking place simultaneously.

Three main intermediate outcomes are distinguished: yields, production costs and prices. Some VSS target yield increases through requirements related to good agricultural practices and the provision of training and/or technical assistance to increase productivity and output. VSS may lower production costs through more efficient input use or lower input prices, improved access to credit or aid support. The VSS' ToCs mostly ignore the possibility of production cost increases related to certification and auditing costs, and to investment and operational costs to comply with VSS requirements. VSS employ a large variety of interventions to support farmgate prices: a price floor, a price premium, good agricultural practices, training and/or technical assistance to increase product quality, market information, access to export and niche markets, or strengthening organizations to increase bargaining power. Besides a focus on yields, production costs and prices, some VSS aim to improve farmers' revenue and welfare directly by providing a social premium. In addition, farmers' adoption of VSS can result in investments in other income-generating activities and spillover effects to non-certified farm output, although these are usually not direct targets of VSS interventions.

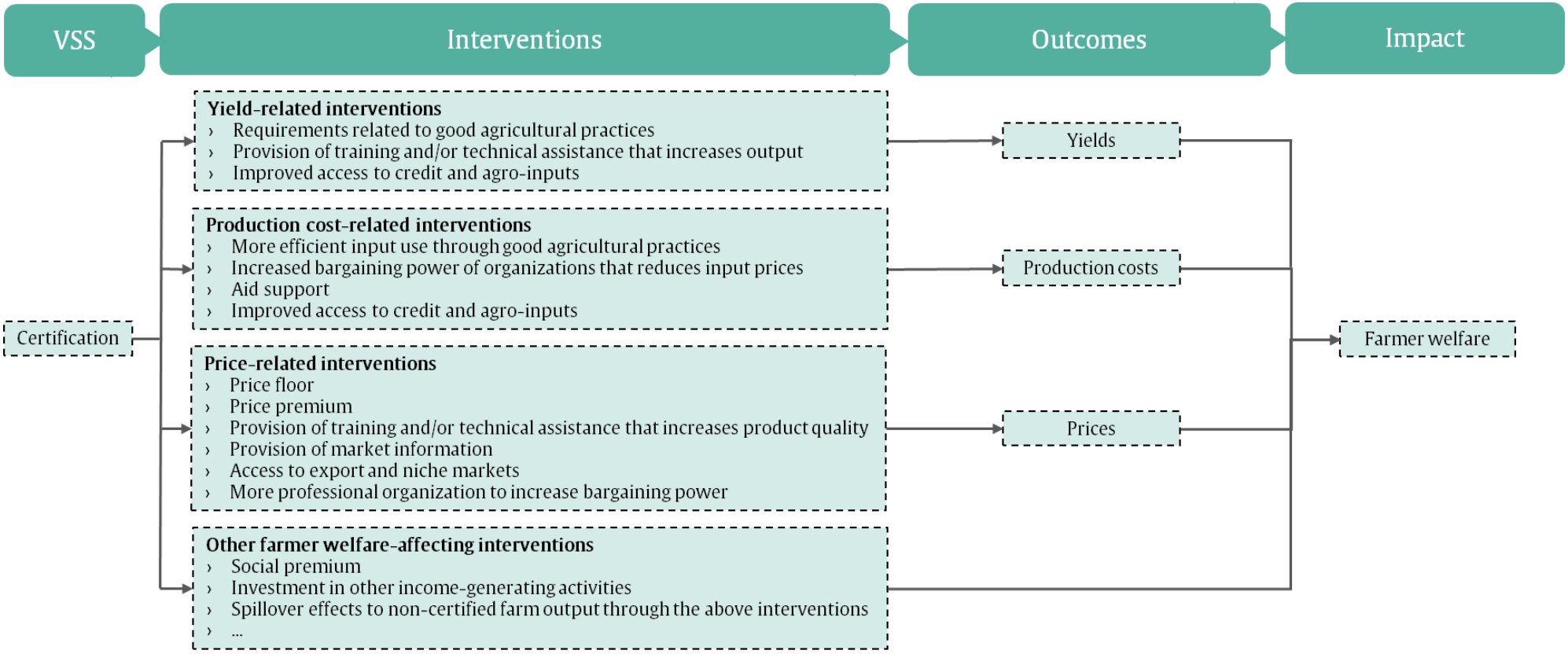


Figure 1: Impact pathways of VSS linking interventions to intermediate outcomes and farmers' welfare as long-term impact based on a review of ToCs of leading VSS

2.2 Literature

We could identify only a handful of theoretical studies on how standards affect the welfare of farmers. The theoretical models developed by Potts (2007), Chen and Lee (2017), Chambolle and Poret (2013), Omidvar and Giannakas (2015), Yu and Bouamra-Mechemache (2016) and Swinnen and co-authors (2015) reveal ambiguous price, yield and cost effects depending on the level of income or development of the country, market conditions, applied technology, commodity characteristics, quality requirements, and the stringency and strength of enforcement of standards.

In contrast, a substantial body of empirical literature exists on the microeconomic welfare impact of VSS in low- and middle-income countries, although with a bias in the evidence towards coffee and Fairtrade certification and towards studies from specific countries in Africa (Traldi, 2021). In a meta-analysis of 97 quantitative studies on the socio-economic impact of smallholder certification, Meemken (2020) concludes that, while on average VSS adoption leads to significant positive (but small) effects on net farm revenue, there is substantial variation in estimated impacts, and that both the specific standard and region matter substantially to explain that variability. Other review studies confirm that VSS adoption has a variable and often weak impact on farmer welfare (Bray and Neilson, 2017; DeFries et al., 2017; Marx et al., 2022; Oya et al., 2018). The following factors are found to facilitate positive welfare effects: effective communication between producers and buyers, proper compliance with VSS, support in the form of pre-financing, long-term contracts, effective horizontal or vertical coordination, social capital generated through horizontal coordination, and assistance in improving product quality (Jena et al., 2012; Ssebunya et al., 2019; Subervie and Vagneron, 2013; Valkila and Nygren, 2010).

Numerous empirical studies have analyzed VSS impacts on yields, prices and costs. In their review, Meemken (2020) concludes that smallholder VSS adoption on average does not change certified crop yields, improves producer prices of certified crops and increases production costs. Yield effects are found to be intervention specific, with VSS focusing on good agricultural practices and providing

training and technical assistance more likely to create positive yield effects and VSS focusing on Organic production practices more likely to create negative yield effects (Akoyi and Maertens, 2018; Ibanez and Blackman, 2016; Mitiku et al., 2017; Vagneron and Roquigny, 2011; Vanderhaegen et al., 2018). Price effects are found to vary by type of price mechanism, with stronger effects for VSS applying a system of quality-based price differentiation (e.g. GlobalGAP) and floor price interventions (e.g. Fairtrade and UTZ) that are particularly effective during commodity price declines (Oya et al., 2018). Also, VSS providing training and price information are more likely to improve producer prices (Jena et al., 2012). Production cost increases mainly relate to higher input and labor costs related to the prohibition of chemical inputs in organic VSS, and to certification costs, although the latter are often paid by donors or exporters (Bachmann, 2012; Beuchelt and Zeller, 2011; Crowder and Reganold, 2015; Holzapfel and Wollni, 2014; Ruben and Zuniga, 2011). Yet, no studies have analyzed how revenue effects are mediated simultaneously through yield, price and cost effects, as we do in this paper.

While empirical studies mostly focus on a single VSS and a single crop in a given setting, some analyze heterogenous effects across countries and regions (Arnould et al., 2009; Qiao et al., 2016), VSS (Akoyi and Maertens, 2018; Chiputwa et al., 2015; Elbers et al., 2015; Haggar et al., 2017; Mitiku et al., 2017; van Rijsbergen et al., 2016; Vanderhaegen et al., 2018), crops (Ruben et al., 2009), or producers (Asfaw et al., 2010; Handschuch et al., 2013; Holzapfel and Wollni, 2014; Meemken, 2021; Subervie and Vagneron, 2013). These studies show that a single VSS can create diverging effects in different settings, that within the same context effects vary across VSS, and that income effects are stronger for larger farms, more educated farmers and farms with shorter physical distance to buyers.

2.3 Hypotheses

Based on the theoretical and empirical evidence above and in reference to the conceptual framework in Figure 1, the general direction and magnitude of the effect of VSS on revenue (i.e., total effect) that

can be explained by effects through yields, costs and prices (i.e., mediation effects[ii]) and the remaining effect (i.e., residual effect) are unclear a priori. We focus on the three main VSS in Peru - Fairtrade, Organic and GlobalGAP - which vary in their focus, requirements and objectives (Table 1). Only Fairtrade has a clear objective in terms of price, yield, production cost and welfare. Fair Trade USA does on the contrary not directly target yield increases but focusses on increasing the total volume of certified sales, rather than individual farm output. Organic and GlobalGAP focus primarily on good agricultural practices (which may affect yields and production costs), and associated quality and price differentiation. Hence, we hypothesize that total and mediation effects vary across VSS. In addition, given the theoretical and empirical evidence presented above, we hypothesize that VSS may have varying effects on different crops.

Table 1: Overview of the objectives of selected VSS as described in their ToCs

VSS	Price objective	Yield objective	Production cost objective	Welfare objective
GlobalGAP version 6	No specific price objective, but reliance on quality-related prices for higher-end markets	/	/	/
Fairtrade International version since 2016-2020[iii]	Minimum price guarantees, the social premium (to invest in economic, social and environmental development), access to pre-finance (credit) and quality improvement	Improved productivity and increasing the total volume of certified sales	Minimum price guarantees and the social premium to ensure the producer organizations receive prices which cover the average cost of sustainable production	Improved incomes, assets and standards of living
Fairtrade USA version 2.1.0	Minimum price guarantees in certain supply chains	Increasing the total volume of certified sales	/	Support sustainable incomes
Organic (IFOAM) version 2019	Reliance on specialty markets with higher prices	/	/	/

[ii] In the mediation literature, mediation and residual effects are also often referred to as indirect and direct effects, respectively.

[iii] Fairtrade International and Fair Trade USA used to form one VSS, but split in 1998 because Fair Trade USA opposed the introduction of a standard for large farms with hired labor.

3 Materials and methods

3.1 Regional context and data

Peru is divided administratively into 1838 districts, 195 provinces and 25 departments, and comprises three agroclimatic regions, from west to east, the coastal zone (*Costa*), the Andean highlands (*Sierra*) and the Amazon rainforest (*Selva*). Fresh fruits and vegetables are mainly grown in the coastal areas, often in agro-industrial companies (of which many are certified), tropical commodities are produced in the Amazon rainforest, mainly by family farms, and staple crops are cultivated in all climate zones. We focus on family farms and the three largest VSS in Peru, including Organic, Fairtrade and GlobalGAP. Organic certification is regulated by the Peruvian Ministry of Agriculture and is the only VSS that has been actively promoted in Peru since 2008. Certification of family farms is often done with the help of buyers or nongovernmental organizations. We focus on coffee and bananas because they have the largest certified area (67% for coffee, 11% for bananas (FAO, 2023; ITC, 2023)), and are cultivated in many departments (Figure A1).

We use secondary data from the *Encuesta Nacional Agropecuaria* (ENA), a nationally representative annual farm survey implemented by the national statistical institutions *Instituto Nacional de Estadística e Informática* and *Dirección Nacional de Censos y Encuestas* (INEI and DNCE, 2023), for the period 2015-2019[iv]. The ENA sample includes approximately 30,000 farms each year, of which about 28,500 are family farms. The sampling strategy is based on the 2012 Agricultural Census as sampling frame and two-stage sampling. In the first stage, 2,086 conglomerates[v] are randomly selected without replacement, and in the second stage, 14 farms are selected randomly without replacement within these conglomerates. The non-response rate is 4% and relates to refusal (0.2%), absence (0.3%) and inactivity of the farm unit (3.5%). The survey data include sampling weights and allow for

[iv] The ENA is also conducted in 2014 but does not include information on farm revenue.

[v] A conglomerate is a georeferenced minimum unit of the 2012 Agricultural Census. In total, there are 19,264 conglomerates in Peru covering on average 66 km² and 120 producers (INEI and DNCE, 2023).

nationally representative statistics and analysis. We remove observations that relate exclusively to livestock production, as these farms differ in several ways from crop-only or mixed farms.

Of the total sample of family farms, 8,943 farms are sampled in at least two of the five years (Table A1), which allows us to apply panel data methods. In addition, we use pseudo-panel methods which permits us to use all available observations from the approximately 28,500 family farms surveyed each year (Deaton, 1985). However, the use of pseudo panels reduces the sample size which, all else equal, leads to less precise results. Given this trade-off and to examine robustness, we report both panel (i.e., farm level) and pseudo-panel (i.e., district level) results[vi].

The ENA data include information on certification at farm-level (not at crop-level) to the following VSS: Organic, Fairtrade, GlobalGAP, Tesco's Nurture add-on to GlobalGAP, Buenas Practicas Manufacturas (i.e., a non-mandatory public standard in Peru) or other VSS that farmers could specify themselves (mentioning Rainforest Alliance, SMETA and various health and safety certifications such as HACCP, BASC, BRC). Given that certification of family farms is relevant only for a handful of crops, we were able to derive crop-specific certification data assuming that the farm-level certificate relates to crops accounting for more than 50% of the farm revenue or harvested area (following Ruben, 2017) or to crops of which more than 50% of the harvest is destined for export, Lima or the agroindustry, where demand for certified produce is high (Schuster and Maertens, 2013; UNFSS, 2022). We verified these assumptions via district-level certification data from Fairtrade International, Fairtrade USA and Organic for 2015-2019, and crop-specific certification data from GlobalGAP for 2021.

[vi] We additionally estimated the pseudo-panel models at the conglomerate and province level, giving very similar results, which are for brevity not reported. Estimations at department level are not feasible due to insufficient observations.

3.2 Econometric approach

To analyze the pathways through which VSS affect net farm revenue, we estimate the following multiple mediation model[vii] at crop-level using structural equation modelling, as suggested by Hoyle (2023) (Eq. 1-2; Figure 2):

$$Net\ revenue_{i,t} = \alpha_0 + \beta_0 VSS_{c,i,t} + \sum_{j=1}^{3} \gamma_{0,j} Mediator_{j,(c,)\,i,t} + \theta_0 X_{i,t} + \mu_{0,c,i} + \tau_{o,t} + \varepsilon_{0,c,\,i,t} \quad (1)$$

$$Mediator_{j,(c,)\,i,t} = \alpha_j + \beta_j VSS_{c,i,t} + \theta_j X_{i,t} + \mu_{j,c,i} + \tau_{j,t} + \varepsilon_{j,c,i,t}\ \forall\ j \in (1,3) \quad (2)$$

where $Net\ revenue_{i,t}$ is the outcome indicator for farm *i* at year *t* as a proxy for farm welfare, calculated as the value of all farm output, certified and non-certified (i.e. crop and livestock products, including self-consumed output) minus production costs (i.e. land lease, deprecation of agricultural equipment and machinery[viii], rental and maintenance of machinery and equipment, farm inputs, fuel, hired labor, services). Data on certification costs are not available. Yet, in Peru these are usually paid by cooperatives, buyers or nongovernmental organization, and not by individual farmers. $VSS_{i,t}$ is the certification status (dummy) of crop *c* of farm *i* at year *t*. $Mediator_{j,(c,)\,i,t}$ represents the revenue-determining factors of interest with subscript *j* taking values from 1 to 3 for the three mediating factors: self-reported average market price (for crop *c* at farm *i* in year *t*), yield (for crop *c* at farm *i* in year *t*) and total production costs (for farm *i* in year *t*). $X_{i,t}$ are time-variant observable farm-level control variables[ix], including total area (levels and squared terms), tropical livestock units[x], the share of irrigated fields, off-farm employment of the household head, and the number of household laborers. $\mu_{c,i}$ are crop- and farm-level fixed effects, capturing time-invariant unobserved heterogeneity. τ_t are

[vii] A multiple mediation model allows to analyze multiple mediators conditionally on the effect of other mediators and compare the magnitudes of these effects (after standardization) while reducing omitted variable bias by analyzing mediation effects conditional on other included mediation pathways (Preacher and Hayes, 2008).

[viii] Since no data on depreciation period or salvage value are available, we assume an average depreciation period of 10 years and a salvage value of 35% of original cost, based on Ellis, Shockley and Shepherd (2021).

[ix] We analyzed the model with 'membership in a farmer organization' as an additional control variable to address potential confounding with the effect of VSS (Table A2). The results remain robust when including this control variable. However, due to possible endogeneity of this variable, we do not include it in the main analyses of the paper.

[x] We use FAO's (2011) and ILRI's (2011) conversion factors.

year dummies, representing time trends and shocks, and $\varepsilon_{i,t}$ is the idiosyncratic error term. We convert all monetary variables to constant 2019 PEN using the consumer price index (The World Bank, 2023). The model is estimated at the crop level, with two of the three mediators varying at the crop level and the third mediator and outcome variable varying at the farm level. This means that we have multiple observations per farm, depending on the number of crops a farm grows. For the pseudo-panel data estimation, we estimate the same model with the district as unit of analysis and with mean aggregated data. To analyze heterogeneity by VSS, we re-estimate the model using a vector of the three main active VSS in Peru (Fairtrade, GlobalGAP and Organic) instead of the variable $VSS_{i,t}$. In doing so, we exclude observations certified to other VSS, as including them would lead to those observations being included in the control group. We do not analyze the effect of multiple certification, due to the small number of multiple certified farms (see Results section). To analyze heterogeneity by crop, we re-estimate the model for the subsamples of coffee and banana farmers.

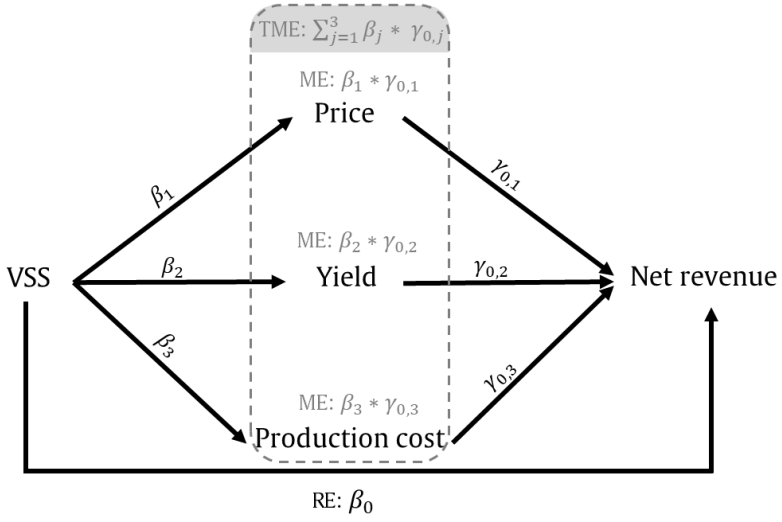


Figure 2: *The estimated multiple mediation model linking VSS to net farm revenue. TME, ME and RE refer to the total mediation, mediation and residual effect, respectively.*

We use Stata's sem package to estimate the multiple mediation model. We apply the full information maximum likelihood (FIML) estimator, because it is the minimum variance unbiased estimator under large sample sizes when the normality assumption is met, can estimate nonlinear models, and is superior in handling missing data (Allison et al., 2017). The FIML estimator assumes missing values

are missing at random and joint normality of all variables. To ensure that the normality assumption is met, we apply the inverse hyperbolic sine transformation[xi] to the mediators and the outcome variable and additionally check the assumption with the Doornik-Hansen test. We allow for non-zero covariance between the mediators to avoid confounding between the effects of the independent variables and the correlation of the residuals. However, since correlation among the mediators might attenuate their estimated coefficients, we contrast the results from the multiple mediation models with estimates from single equation models (Table A3) and from single mediation models (Table A4). We analyze the goodness of fit using chi-square, root mean square error of approximation and the comparative fit index based on the cutoffs of Hu and Bentler (1999), which indicate good model fit, and therefore not reported for brevity. We report both unstandardized and standardized coefficients. The latter allows to compare the relative magnitudes of the estimates.

Based on the estimation results, we calculate the effect of VSS on revenue (i.e., total effect) that can be explained by effects through the mediators (i.e., specific mediation effect) and through all mediators together (i.e., total mediation effect) using Sobel's method with bootstrapped standard errors (200 replications), as recommended by Preacher and Hayes (2008), and the Holm-Bonferroni correction to account for multiple hypothesis testing (Holm, 1979) (Eq. 3-4; Figure 2):

$$Specific\ mediation\ effect = \beta_j * \gamma_{0,j}\ \forall\ j \in (1,3) \tag{3}$$

$$Total\ mediation\ effect = \sum_{j=1}^{3} \beta_j * \gamma_{0,j} \tag{4}$$

Our basic estimation of the mediation model includes fixed effects (FE) to account for time-constant unobserved heterogeneity. However, potential endogeneity may also result from (1) reverse causality,

[xi] The inverse hyperbolic sine transformation is defined for negative and zero earnings, unlike the logarithmic transformation (Bellemare and Wichman, 2020).

(2) systematic measurement error[xii], (3) time-varying omitted variables and (4) attrition. To account for (1), we compare the FE estimates with a first difference model[xiii] (FD) (Table A5). Reverse causality may violate the strict exogeneity assumption of the FE model, but not the sequential exogeneity assumption of the FD model, which, however, is known to be less efficient when the number of time periods exceeds two (Wooldridge, 2010). To account for (2), we control for sources of time-invariant systematic measurement error by unit and time fixed effects. In addition, we conduct a robustness check by reevaluating the mediation model excluding outliers (i.e. z-score above 3) and using median aggregation, as recommended by Bellemare and co-authors (2018). Classical measurement error[xiv], causing attenuation bias, could remain but is reduced by using pseudo-panel methods. To account for (3), we apply an instrumental variable method to each mediation model equation as a robustness check (Table A6). Since no valid external instruments are found, we rely on Lewbel's method (2012) which creates internal instruments by exploiting information contained in the heteroscedasticity of the first-stage residuals. The Lewbel estimator requires two additional assumptions. The first is $Cov(Z, \varepsilon_1^2) \neq 0$, meaning that the first stage residuals are heteroscedastic, which we check using the Breusch-Pagan test. The second is $Cov(Z, \varepsilon_1 \varepsilon_2) = 0$, meaning that the instrument is valid because it is uncorrelated with the second-stage residuals, which we evaluate with both the Pagan-Hall test and Hansen's J statistic. In addition, we check for weak instruments with the Kleibergen-Paap rk Wald F statistic, under-identification with the Kleibergen-Paap rk LM statistic and endogeneity with the difference of two Hansen's J statistics. We estimate Lewbel's method using the generalized method of moments estimator. To account for (4), we recalculate the FE-FIML estimates using a balanced panel for 2016-2019.

[xii] This can be an issue, especially with Fairtrade price data, as cooperatives often pay farmers partially in advance with an end-of-year payment or use part of the social premium to give additional payments to farmers which may or may not be perceived by farmers as the price that they receive for their products.

[xiii] The FD estimations are calculated using the standard maximum likelihood estimator and not the FIML estimator since the latter treats the differenced values as missing in this case.

[xiv] This can be especially the case with the recall of price data that varies by batch, time of sale, and especially by quality.

4 Results

4.1 Descriptive statistics

Table 2 provides an overview of the sample. The number of certified farms is relatively small, but comparable to the sample sizes of surveys specifically designed to analyze the impact effectiveness of VSS (Meemken, 2020). Looking at the types of VSS (Figure 3), Organic is the most prevalent, followed by GlobalGAP and Fairtrade. VSS are widespread across the country (Figure A2). Organic and Fairtrade certification are especially common in the north, while GlobalGAP is also common at the coast. Only 16 to 25% of certified farms are multiple certified over the different years, documenting a rather low incidence of multiple certification.

Table 2: Overview of the sample and sub-samples of certified farms, by year and administrative level

	2015		**2016**		**2017**		**2018**		**2019**	
N	**Total**	**Certified**	**Total**	**Certified**	**Total**	**Certified**	**Total**	**Certified**	**Total**	**Certified**
Farms	28,989	476	29,580	438	29,610	436	29,624	456	29,555	425
Family farms										
Farms	26,684	215	27,219	239	27,254	220	27,203	204	27,300	165
Conglomerates	2067	101	2069	90	2068	93	2069	77	2078	83
Districts	1025	75	1023	61	977	63	968	57	985	61
Provinces	192	47	192	38	193	38	192	36	194	38
Departments	24	21	24	16	24	18	25	16	25	20

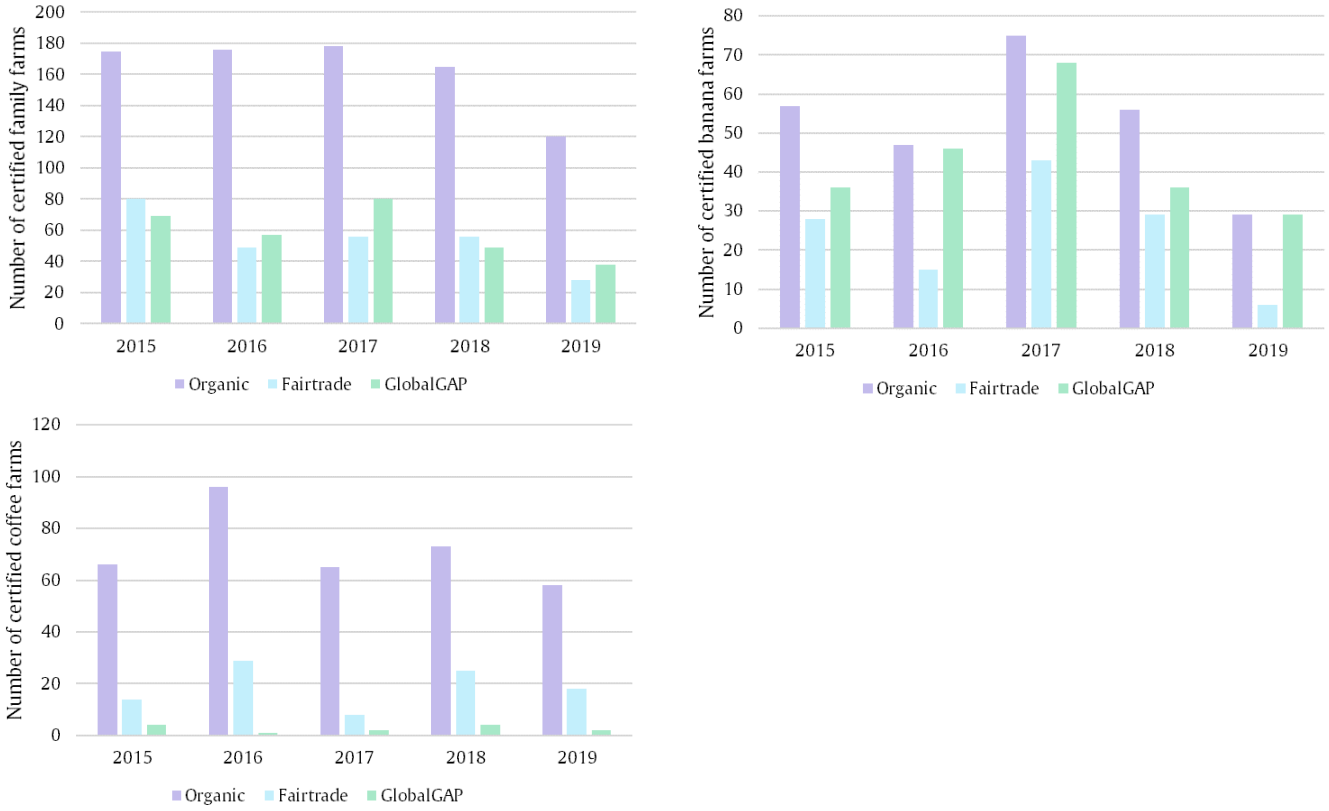


Figure 3: Overview of the VSS adopted by family farms in the ENA sample for (a) all family farms, (b) family farms certified for banana and (c) family farms certified for coffee

Comparing certified with non-certified coffee and banana farms, we find that certified farms have higher market prices and yields (Table 3). On average, certified farms also have higher production costs and net farm revenues (Table 4). In addition, certified farms maintain fewer livestock but cultivate almost twice as much land, documenting that the smallest farms are less involved in certification. Almost all certified farms are member of a producer organization while hardly any of the non-certified farms is. Other differences between certified and non-certified farmers relate to family size, gender, language, ethnicity, education and off-farm employment.

Table 3: Descriptive statistics of the total sample of family farms by crop and certification status of the crop

	Coffee			Bananas		
	Non-certified	**Certified**		**Non-certified**	**Certified**	
N	10,556	387		27,600	404	
Market price (2019 PEN/kg)	6.26	7.25	***	0.71	1.05	***
	(0.07)	(0.12)		(0.02)	(0.02)	
Yield (kg/ha)	679.07	822.38	***	9,564	29,610	***
	(16.90)	(56.40)		(183)	(1,791)	

*Note: Sampling weights used for all statistics. Standard errors in parentheses. Comparison between the certified and non-certified subgroups is based on a Mann-Whitney test. Significant differences of the mean of the subgroups are shown with * p < 0.1, ** p < 0.05, *** p < 0.01.*

Table 4: Descriptive statistics of the total sample of family farms by certification status

	Non-certified	**Certified**	
N	135,660	1,043	
Annual net farm income (2019 PEN)	5,869 (194.5)	23,379 (2,153)	***
Total farm production costs (2019 PEN)	4,202 (91.1)	7,661 (773)	***
Experience in independent agricultural activity (years)	26.24 (0.10)	23.98 (0.59)	
Tropical livestock units	3.48 (0.08)	1.75 (0.18)	***
Use of advanced irrigation techniques (% of area)	0.38 (0.01)	0.45 (0.07)	
Total harvested area (ha)	1.66 (0.05)	3.22 (0.31)	***
Total area (ha)	5.63 (0.22)	5.96 (0.68)	***
Self-reported distance to the capital district (hours)	1.53 (0.03)	1.24 (0.16)	
Household size	3.55 (0.01)	3.95 (0.08)	***
Number of household laborers on the farm	2.32 (0.01)	2.31 (0.06)	***
Age of the household head (years)	53.11 (0.11)	50.79 (0.70)	
Part of a farmers' association/cooperative/committee (0/1)	0.05	0.93	***
Female-headed household (0/1)	0.30	0.15	***
Secondary or tertiary education of the household head (0/1)	0.33	0.48	***
Spanish as first language of the household head (0/1)	0.57	0.87	***
Indigenous ethnic group of the household head (0/1)	0.53	0.20	***
Off-farm employment by household head (0/1)	0.52	0.48	**

*Note: Sampling weights used for all statistics. Standard errors for non-binary variables in parentheses. Comparison between the certified and non-certified subgroups is based on a Mann-Whitney or Chi² test. Significant differences of the mean of the subgroups are shown with * p < 0.1, ** p < 0.05, *** p < 0.01.*

4.2 Results of multiple mediation analysis

Figure 4 visualizes the standardized estimates of (a) the single equation model and (b-d) the mediation models at the farm level. Table A3 presents the unstandardized regression results for the single equation models and Tables 5-7 for the mediation models at the farm and district level. Overall, at farm level, we find that the total effect of VSS on net farm revenue is small and insignificant for all FE estimators, indicating that VSS do not improve farm revenue (Figure 4a; Table A3). In the mediation model, we explicitly decompose this total effect into mediation effects and the residual effect. We find a negative total mediation effect offsetting a positive residual effect, explaining the insignificant total effect (Figure 4b; Table 5). This implies that VSS affect farm revenue negatively through prices, yields and costs jointly and positively through other mechanisms.

What are the main determinants contributing to this negative total mediation effect? We find that production costs, followed by price, are the main mediators, while yield does not contribute (Figure 4b; Table 5). In real terms, VSS increase prices by 7%, which in turn increase net farm revenue by 23%, but also increase production costs by 52%, which reduce net farm income by 58%[xv]. The last rows of Table 5 report a significant covariance between VSS and certain time-constant unobservable factors, demonstrating the need to account for unobserved heterogeneity using unit and time fixed effects, as we do in our models.

At the district level, the results slightly deviate: here we find that both the total mediation effect and the residual effect are insignificant (Table 5). The former is due to a larger mediation effect through prices that offsets the negative mediation effect through production costs: an increasing share of certified farmers in the district raises average prices by 23%, thereby increasing net farm revenue by 53%. The latter implies that the mediators explain most of the effect on net farm revenue at the district level, presumably due to averaging out the farm-level idiosyncratic shocks.

[xv] Calculated based on the unstandardized estimates using the formula provided by Bellemare and Wichman (2020).

Robustness checks show that single equation model results are similar (Table A3), but these do not allow us to compare the relative sizes of the effects. The results of the single mediation models also demonstrate comparable results (Table A4), indicating that the correlation between the mediators does not attenuate their estimated coefficients. In addition, the results of the FD-FIML estimator are comparable (Table A5), implying that a feedback mechanism is negligible. Furthermore, the results of the Lewbel IV estimation show similar effects but larger coefficients than the FE-FIML results (Table A6). The Kleibergen-Paap rk Wald F statistic, with critical values between 8.52 and 11.53 for 10% relative bias[xvi], suggests that some FE-IV estimates may suffer from a weak instrument problem. However, we find that, except for net farm income, the effects become smaller at the district level, so the estimates become more precise when fewer idiosyncratic shocks can affect the relationship, suggesting that the fixed effects already account for most of the unobserved heterogeneity. Additional robustness checks excluding outliers, using median aggregation and using a balanced panel provide similar results, indicating that outliers and attrition are not influential, and are therefore omitted for brevity.

[xvi] We follow the method proposed by Baum et al. (2007) to apply Stock and Yogo's critical values initially established for the Cragg-Donald statistic, since no critical values exist for the Kleibergen-Paap statistic.

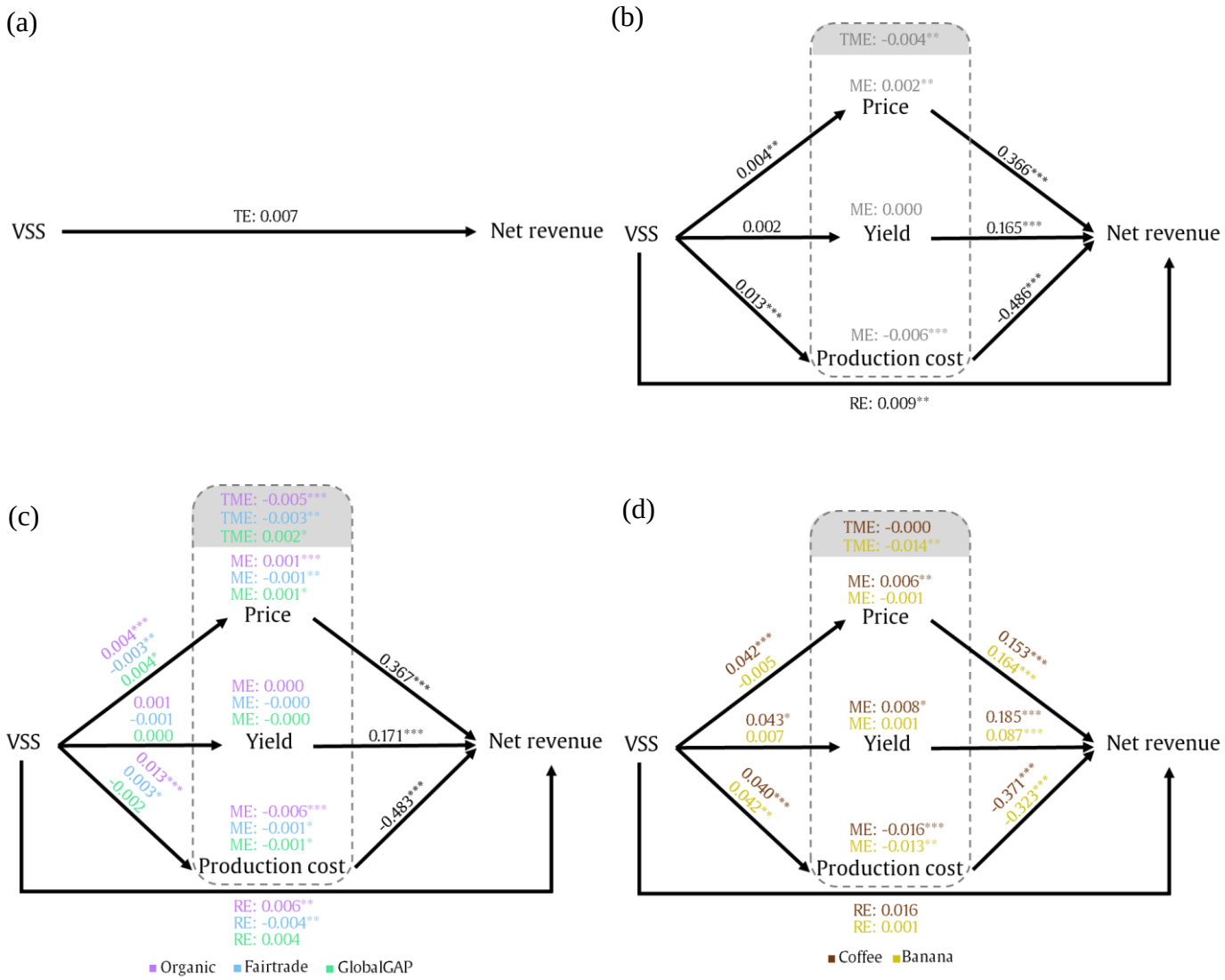


Figure 4: Standardized results from the multiple mediation models at farm level for (a) the singe equation model, (b) all crops, (c) all crops and standard-specific dummies and (d) for the subsamples of coffee and banana farmers. Control variables, crop-, farm- and year- fixed effects are included but estimated coefficients not shown for brevity. TE stands for total effect, TME stands from total mediation effect, ME for mediation effect and RE for residual effect. Unstandardized results are presented in Tables 5-7.

Table 5: Unstandardized results of the mediation model for all crops and all VSS using the FE-FIML estimator

Spatial level	**Farm**	**District**
Dependent variable: Net farm revenue (2019 PEN)		
Certification (β_0)	1.167 (0.465)**	0.079 (1.518)
Market price ($\gamma_{0,1}$)	3.163 (0.273)***	2.092 (0.214)***
Yield ($\gamma_{0,2}$)	0.928 (0.072)***	0.640 (0.069)***
Production cost ($\gamma_{0,3}$)	-1.965 (0.034)***	-1.852 (0.034)***
Dependent variable: Market price (2019 PEN/kg)		
Certification (β_1)	0.067 (0.031)**	0.210 (0.073)***
Dependent variable: Yield (kg/ha)		
Certification (β_2)	0.043 (0.044)	-0.035 (0.074)
Dependent variable: Production cost (2019 PEN)		
Certification (β_3)	0.423 (0.115)***	0.471 (0.259)*
Mediation effects		
Market price ($\beta_1 * \gamma_{0,1}$)	0.212 (0.100)**	0.439 (0.160)***
Yield ($\beta_2 * \gamma_{0,2}$)	0.040 (0.041)	-0.022 (0.048)
Production cost ($\beta_3 * \gamma_{0,3}$)	-0.831 (0.225)***	-0.873 (0.481)*
Total mediation effect ($\sum_{j=1}^{3} \beta_j * \gamma_{0,j}$)	-0.578 (0.262)**	-0.456 (0.494)
N crops	49,718	44,948
Doornik-Hansen test (p-value)	0.000	0.000
F test (p-value)	0.000	0.000
Hausman test (p-value)	0.000	0.000
Cov(Certification, $\mu_{0,c,i}$)	0.007 (0.002)***	0.003 (0.001)***
Cov(Certification, $\mu_{1,c,i}$)	0.001 (0.000)**	0.001 (0.000)***
Cov(Certification, $\mu_{2,c,i}$)	0.001 (0.000)	-0.000 (0.000)
Cov(Certification, $\mu_{3,c,i}$)	0.000 (0.000)	0.000 (0.000)

*Note: *p<0.1; ** p<0.05; *** p<0.01. Clustered standard errors for the FE-FIML estimates and bootstrapped standard error for the mediation effect estimates in parentheses. We report the Doornik-Hansen test for multivariate normality, F test for unobserved heterogeneity, Hausman test for comparison with a random effects model and covariance between VSS and certain time-constant unobservable factors. Control variables, crop-, farm- and year- fixed effects are included but estimated coefficients not shown for brevity. Full regression results can be found in Table A2.*

4.3 Heterogeneity by VSS and by crop

Do the results differ by VSS? At the farm level, we find that the effects for Organic and GlobalGAP are quantitatively similar to the overall effects discussed in Section 4.2 (Figure 4c; Table 6). Notable exceptions are that the total mediation effect of Organic is larger in magnitude due to a larger mediation effect through costs (65%), and that of GlobalGAP is positive due to a smaller mediation effect through costs (22%). In addition, the residual effect of Organic is slightly smaller in magnitude and that of GlobalGAP is insignificant. Yet, the effects for Fairtrade VSS deviate substantially. Fairtrade has a larger negative total mediation effect and negative residual effect, implying that Fairtrade negatively affects net revenue. The larger negative total mediation effect is the result of a smaller mediation effect through production costs (51%) and, surprisingly, a negative mediation effect through price (35%), implying that Fairtrade reduces farm revenues through lower prices compared to non-certified farms.

At the district level, we find comparable results for Fairtrade, but some differences for Organic and GlobalGAP (Table 6). An increasing share of Organic certified farmers in the district leads to a positive total mediation effect due to a smaller mediation effect through costs (5%), while an increasing share of GlobalGAP certified farmers in the district leads to a negative total mediation effect due to a larger mediation effect through costs (93%).

Table 6: Unstandardized results of the mediation model for all crops and standard-specific dummies using the FE-FIML estimator

Spatial level	**Farm**	**District**
Dependent variable: Net farm revenue (2019 PEN)		
Organic (β_0)	0.932 (0.473)**	1.166 (1.360)
Fairtrade (β_0)	-1.759 (0.720)**	-4.843 (3.506)
GlobalGAP (β_0)	0.837 (0.754)	5.751 (3.658)
Market price ($\gamma_{0,1}$)	3.194 (0.246)***	2.102 (0.215)***
Yield ($\gamma_{0,2}$)	0.976 (0.104)***	0.639 (0.070)***
Production cost ($\gamma_{0,3}$)	-1.969 (0.038)***	-1.853 (0.034)***
Dependent variable: Market price (2019 PEN/kg)		
Organic (β_1)	0.071 (0.026)***	0.187 (0.118)*
Fairtrade (β_1)	-0.130 (0.040)***	-0.381 (0.151)**
GlobalGAP (β_1)	0.089 (0.053)*	0.350 (0.150)**
Dependent variable: Yield (kg/ha)		
Organic (β_2)	0.028 (0.042)	-0.069 (0.088)
Fairtrade (β_2)	-0.107 (0.074)	0.204 (0.188)
GlobalGAP (β_2)	-0.015 (0.054)	0.011 (0.132)
Dependent variable: Production cost (2019 PEN)		
Organic (β_3)	0.524 (0.107)***	-0.085 (0.052)*
Fairtrade (β_3)	0.328 (0.184)*	0.164 (0.099)*
GlobalGAP (β_3)	-0.122 (0.180)	1.033 (0.920)
Mediation effects Organic		
Market price ($\beta_1 * \gamma_{0,1}$)	0.225 (0.085)***	0.393 (0.238)*
Yield ($\beta_2 * \gamma_{0,2}$)	0.027 (0.041)	-0.044 (0.057)
Production cost ($\beta_3 * \gamma_{0,3}$)	-1.032 (0.210)***	0.157 (0.095)*
Total mediation effect Organic ($\sum_{j=1}^{3} \beta_j * \gamma_{0,j}$)	-0.780 (0.233)***	0.507 (0.307)*
Mediation effects Fairtrade		
Market price ($\beta_1 * \gamma_{0,1}$)	-0.416 (0.133)***	-0.802 (0.328)**
Yield ($\beta_2 * \gamma_{0,2}$)	-0.104 (0.073)	0.130 (0.121)
Production cost ($\beta_3 * \gamma_{0,3}$)	-0.646 (0.362)*	-0.303 (0.184)*
Total mediation effect Fairtrade ($\sum_{j=1}^{3} \beta_j * \gamma_{0,j}$)	-1.167 (0.410)***	-0.974 (0,590)*
Mediation effects GlobalGAP		
Market price ($\beta_1 * \gamma_{0,1}$)	0.283 (0.172)*	0.736 (0.320)**
Yield ($\beta_2 * \gamma_{0,2}$)	-0.014 (0.053)	0.007 (0.085)
Production cost ($\beta_3 * \gamma_{0,3}$)	-0.241 (0.152)*	-1.914 (1.705)
Total mediation effect GlobalGAP ($\sum_{j=1}^{3} \beta_j * \gamma_{0,j}$)	0.510 (0.303)*	-1.171 (1.823)
N crops	49,718	44,948
Doornik-Hansen test (p-value)	0.000	0.000

F test (p-value)	0.000	0.000
Hausman test (p-value)	0.000	0.000

*Note: *p<0.1; ** p<0.05; *** p<0.01. Clustered standard errors for the FE-FIML estimates and bootstrapped standard error for the mediation effect estimates in parentheses. We report the Doornik-Hansen test for multivariate normality, F test for unobserved heterogeneity and the Hausman test for comparison with a random effects model. The covariances between the three VSS and the unobserved heterogeneity are not shown here for brevity. Control variables, crop-, farm- and year- fixed effects are included but estimated coefficients not shown for brevity.*

Do the results differ by crop? At the farm level, compared to the results in Section 4.2, coffee certification results in an insignificant total mediation effect due to a smaller mediation effect through costs (42%) and a small positive mediation effect through yields (29%). Banana certification leads to a similar total mediation effect as estimated in Section 4.2, except that the mediation effect through production costs is smaller (44%). In addition, both coffee and banana VSS have insignificant residual effects on net revenues. At the district level, the results are largely similar, except that the mediation effect through yields for certified coffee farmers and the mediation effect through production costs for certified banana farmers are insignificant.

Table 7: Unstandardized results of the mediation model for the subsamples of coffee and banana farmers respectively and all VSS using the FE-FIML estimator

Crop	**Coffee**		**Banana**	
Spatial level	**Farm**	**District**	**Farm**	**District**
Dependent variable: Net farm revenue (2019 PEN)				
Certification (β_0)	0.543 (0.979)	0.885 (3.017)	0.058 (0.362)	-0.013 (1.175)
Market price ($\gamma_{0,1}$)	3.102 (0.916)***	1.557 (1.366)	1.727 (0.234)***	1.012 (0.597)*
Yield ($\gamma_{0,2}$)	1.740 (0.336)***	0.365 (0.690)	0.959 (0.234)***	1.075 (0.456)**
Production cost ($\gamma_{0,3}$)	-1.118 (0.147)***	-1.751 (0.331)***	-1.013 (0.084)***	-1.411 (0.226)***
Dependent variable: Market price (2019 PEN/kg)				
Certification (β_1)	0.070 (0.014)***	0.074 (0.045)*	-0.018 (0.050)	0.169 (0.070)**
Dependent variable: Yield (kg/ha)				
Certification (β_2)	0.154 (0.087)*	0.252 (0.153)*	0.023 (0.059)	-0.055 (0.140)
Dependent variable: Production cost (2019 PEN)				
Certification (β_3)	0.441 (0.163)***	0.408 (0.247)*	0.523 (0.217)**	-0.090 (0.513)
Mediation effects				
Market price ($\beta_1 * \gamma_{0,1}$)	0.219 (0.079)***	0.116 (0.085)*	-0.030 (0.086)	0.171 (0.104)*
Yield ($\beta_2 * \gamma_{0,2}$)	0.269 (0.159)*	0.092 (0.156)	0.022 (0.057)	-0.060 (0.153)
Production cost ($\beta_3 * \gamma_{0,3}$)	-0.493 (0.192)***	-0.714 (0.355)**	-0.530 (0.223)**	0.128 (0.724)
Total mediation effect ($\sum_{j=1}^{3} \beta_j * \gamma_{0,j}$)	-0.005 (0.242)	-0.507 (0.308)	-0.538 (0.248)**	0.239 (0.726)
N crops	609	264	2,225	621
Doornik-Hansen test (p-value)	0.000	0.000	0.000	0.000
F test (p-value)	0.000	0.000	0.000	0.000
Hausman test (p-value)	0.000	0.000	0.000	0.000
Cov(Certification, $\mu_{0,c,i}$)	0.008 (0.019)	0.014 (0.017)	0.045 (0.013)***	0.025 (0.014)*
Cov(Certification, $\mu_{1,c,i}$)	0.003 (0.001)**	0.003 (0.001)***	0.004 (0.001)***	-0.001 (0.001)**
Cov(Certification, $\mu_{2,c,i}$)	0.001 (0.003)	-0.000 (0.002)	0.020 (0.004)***	0.006 (0.002)**
Cov(Certification, $\mu_{3,c,i}$)	0.009 (0.008)	0.000 (0.004)	0.010 (0.004)**	0.008 (0.004)*

*Note: *p<0.1; ** p<0.05; *** p<0.01. Clustered standard errors for the FE-FIML estimates and bootstrapped standard error for the mediation effect estimates in parentheses. We report the Doornik-Hansen test for multivariate normality, F test for unobserved heterogeneity, Hausman test for comparison with a random effects model and covariance between VSS and certain time-constant unobservable factors. Control variables, crop-, farm- and year- fixed effects are included but estimated coefficients not shown for brevity.*

5 Discussion

We find that VSS adoption in Peru generally does not affect yield while it affects farm revenue positively through higher prices and negatively through higher production costs. Yet, the positive price effect is not large enough to compensate for the higher production costs and improve farm revenue. This contradicts the average positive effect of VSS adoption on farm revenue of 52% that Meemken (2020) finds in a meta-analysis of case-studies in various countries, but is in line with previous findings from Peru (Ruben and Fort, 2012). Our finding that price, rather than yield, is the main channel through which VSS improve farm revenue is important. The channels of effects have been discussed to some extent in the literature based on separate single equation estimates of revenue, yield, price and production cost effects without firm conclusions (e.g. in Estrella et al., 2022; Qiao et al., 2016; Vanderhaegen et al., 2018). We believe our estimates of farm-level VSS impacts, in comparison to those of other studies, are less prone to overestimation because deliberate panel data methods are used, have larger external validity because nationally representative data and multiple years of observations are used, and better document channels of effects because mediation effects are explicitly considered.

The positive price effect we find holds for GlobalGAP and Organic, but not for Fairtrade, where we find a negative price effect. The strong positive price effect for GlobalGAP and Organic points to a higher effectiveness of a quality-based price differentiation system than price premia or a price floor as applied by Fairtrade, as already indicated for GlobalGAP (Oya et al., 2018). The lower effectiveness of Fairtrade's price mechanism can be explained by observations from stakeholder interviews conducted in April-May 2022, and from other studies. Only a fraction of a Fairtrade-certified crop is sold as certified produce due to either a low demand or limited processing capacity. This results in certification rents to be eliminated when distributed to a larger number of producers (de Janvry et al., 2015). In addition, the cooperative certification system of Fairtrade sometimes results in a lack of farmer awareness about their certification status, diminishing their bargaining power to receive the price premium from the cooperative (Jena et al., 2012; Snider et al., 2017), in self-reported producer prices being lower than prevailing market prices because of cooperative deductions to cover

operational costs (Mitiku et al., 2017) or because of advances and payments at the end of the year or monetary benefits of the social premium not factored into the self-reported producer prices. Yet, we need to note that commodity prices were rather high during the study period, rendering the Fairtrade minimum price irrelevant (Snider et al., 2017). The positive price effect we find holds also for coffee but not for banana at the farm level, which might relate to differences in type of commodity (higher value-added horticulture vs traditional intermediate products) and the associated value chains, but also to differences in prevalence of VSS, with a larger share of Fairtrade certified banana farmers occurring in the panel.

Furthermore, the large production cost effect and non-significant yield effect are striking, because VSS are supposed to improve farming systems, resulting primarily in better yields, be it at higher production costs. The finding holds for all three VSS and crops, except coffee which has a small positive yield effect at the farm level. The large production cost effect is due to both higher labor and input costs, which is in line with previous case-study results (Bachmann, 2012; Beuchelt and Zeller, 2011; Crowder and Reganold, 2015; Ruben and Zuniga, 2011). The non-significant yield effect is in line with previous studies on non-Organic certified farms, but contradicts findings of on average lower yields for Organic certified farms in low- and middle-income countries (Meemken, 2020). However, based on our results, we calculate that there is a large potential yield effect: if yield were to increase by 5%, this would result in a 4.6% increase in net revenue. An increase in producer prices by 5%, however, would give rise to a larger effect on net revenue, namely by 14.1%.

Overall, we find similar effects for Organic and GlobalGAP, different effects for Fairtrade, and differences in effects by crop. The three studied VSS differ in their standard design and in their occurrence in distinct types of commodities and associated value chains (i.e., Organic and Fairtrade mainly for traditional intermediate products vs GlobalGAP for higher value-added horticultural products). This might indicate that both the type of commodity and associated supply chains, and the standard design matters importantly for the welfare impact of VSS and the impact pathway, as suggested by Meemken (2020) and Oya and co-authors (2018).

Finally, we find that a significantly positive residual effect exists at the farm level, implying that there are other important mechanisms through which VSS increase farm revenue. These could relate, for example, to the social premium provided by Fairtrade (Sellare, 2022) or to spillover effects on non-certified farm output (Wollni and Andersson, 2014), to effects of membership to a farmer organization (Sellare et al., 2020) or to time-varying omitted variables that we could not take into account. This stresses the importance of considering non-certified farm output and non-farm revenue to understand the welfare effects of VSS.

Our appraisal shows that mediation analyses can be useful to unravel *how* VSS affect farmers. Several limitations of the present study should be acknowledged. A first limitation is the lack of strong external instruments to fully control for time-varying unobserved heterogeneity in the mediation model, such as dynamics in structures and services of organizational membership, as pointed out by Sellare (Sellare, 2022). To fully address selection bias relevant and valid external instruments for both the treatment and mediators are needed. Other limitations relate to data availability. Since the dataset does not include data on the value of off-farm employment, we focus on net farm revenue rather than total household income as outcome indicator. Yet, off-farm employment can be an important source of income for farm-households, and may be affected by VSS adoption. In addition, the extent to which we can disentangle the effect by VSS, and crop is limited given small sub-sample sizes. Moreover, crop-specific certification data are derived using assumptions verified with data from certification programs.

6 Policy implications

Our results have three important policy implications. First, we show that VSS in Peru generally neither improve nor worsen farmer welfare. So where is the social welfare gain? Most VSS contain multiple requirements related to social, economic and environmental sustainability. Studies show that VSS are generally more successful at creating environmental benefits than socioeconomic benefits (Traldi, 2021). This suggests that the potential environmental benefits do not come at the expense of farmer

welfare, but that the costs are borne by the system. In addition, our finding that there is no welfare gain for farmers in Peru puts doubt on the effectiveness of VSS as market-based instrument to improve economic sustainability in global value chains. VSS implicitly shift responsibility for sustainability management from the public to the private sector, and explicitly so through the recent integration of VSS in trade agreements and due diligence requirements. This raises the question whether more government-regulated trade or other market-based tools would be a better way to ensure economic sustainability in Peru. Compared to other regions, studies suggest that VSS produce more favorable results in Africa, especially in reducing production costs (Meemken, 2020). This may be related to how value chains are organized in different contexts (Otsuka and Fan, 2021) and is therefore important to consider in policies regarding VSS.

Second, we find that relying on price premiums or floor prices to improve farmer welfare can be problematic. A quality-based price differentiation system (e.g., in Organic and GlobalGAP) is more effective than a floor price or price premium (e.g. in Fairtrade), likely because the former is more consistent with farmers' economic incentives. This highlights the need for VSS to develop requirements and approaches that are incentive compatible. This is important because studies have shown that price is a key driver of farmer compliance, especially in low- and middle-income countries (UNFSS, 2022), and because we find that price (rather than yield) is the main channel through which VSS affect farmer welfare, with a 5% price increase creating a 14% revenue increase. Quality-based price differentiation systems may suffer less from problems of oversupply and elite capture of price premiums, which are reported for Fairtrade supply chains.

Third, we find no yield improvements from certification although production costs are increased, which is worrisome because most family farms in low- and middle-income countries have large yield gaps relative to theoretically attainable yields and relative to the rest of the world (Cassman and Grassini, 2020). However, we find that there is a large potential for VSS to contribute to more productive agri-food systems: a 5% increase in yields would result in an increase in net revenue by 4.6%. Realizing the yield potential could be a strategy to partially offset the higher production costs

related to certification and would pay off more to farm revenue if done simultaneously with an effort to improve producer prices. In addition, realizing the yield potential could be compatible with environmental goals of VSS. We believe that in standard design, more attention should be paid to increasing individual farm productivity than to increasing the market share of certified products which could better reflect farmers' economic incentives to improve farm income. This, however, requires consideration of local agroclimatic conditions and soil fertility, which is less compatible with the idea of ‘standardizing’ production practices in VSS. In addition, our results confirm that increasing farmer welfare through VSS requires attention to reducing investment and operational costs related to certification, for example through technical producer support, harmonization and mutual recognition of VSS or microfinance where compliance with VSS can act as a good-payment signaling mechanism (UNFSS, 2022).

7 Conclusion

How VSS affects farmer welfare has been called into question. Using nationally representative data from Peru over several years, crops and VSS, and multiple mediation modeling, we show that prices (and not yields) are the main channel through which VSS affect farmer welfare in Peru. However, higher prices cannot offset higher production costs, resulting in zero net welfare gains. In addition, we find that certification does not affect farmer welfare through yield, but that yield improvements have a large potential for VSS to improve farm revenue. We find heterogeneity in the effects by standard and crop, with crops certified to standards that apply a system of quality-based price differentiation having the largest impact on farmer welfare through a price effect. For VSS to effectively improve economic sustainability, policy attention is needed to improvements in standard design. Key areas for improvement include making land productivity an explicit objective, incorporating incentive-compatible practices and pricing schemes into the standards, and reducing implementation costs.

References

Akoyi, K.T., Maertens, M., 2018. Walk the Talk: Private Sustainability Standards in the Ugandan Coffee Sector. J. Dev. Stud. 54, 1792–1818. https://doi.org/10.1080/00220388.2017.1327663

Allison, P.D., Williams, R., Moral-Benito, E., 2017. Maximum Likelihood for Cross-lagged Panel Models with Fixed Effects: Socius 3. https://doi.org/10.1177/2378023117710578

Arnould, E.J., Plastina, A., Ball, D., 2009. Does Fair Trade Deliver on Its Core Value Proposition? Effects on Income, Educational Attainment, and Health in Three Countries. J. Public Policy Mark. 28, 186–201. https://doi.org/10.1509/JPPM.28.2.186

Asfaw, S., Mithöfer, D., Waibel, H., 2010. Agrifood supply chain, private-sector standards, and farmers' health: evidence from Kenya. Agric. Econ. 41, 251–263. https://doi.org/10.1111/J.1574-0862.2010.00443.X

Bachmann, F., 2012. Potential and limitations of organic and fair trade cotton for improving livelihoods of smallholders: Evidence from Central Asia. Renew. Agric. Food Syst. 27, 138–147. https://doi.org/10.1017/S1742170511000202

Barham, B.L., Callenes, M., Gitter, S., Lewis, J., Weber, J., 2011. Fair Trade/Organic Coffee, Rural Livelihoods, and the "Agrarian Question": Southern Mexican Coffee Families in Transition. World Dev. 39, 134–145. https://doi.org/10.1016/J.WORLDDEV.2010.08.005

Barham, B.L., Weber, J.G., 2012. The Economic Sustainability of Certified Coffee: Recent Evidence from Mexico and Peru. World Dev. 40, 1269–1279. https://doi.org/10.1016/J.WORLDDEV.2011.11.005

Baum, C.F., Schaffer, M.E., Stillman, S., 2007. Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing. Stata J. 7, 465–506. https://doi.org/10.1177/1536867X0800700402

Bellemare, M.F., Fajardo-Gonzalez, J., Gitter, S.R., 2018. Foods and fads: The welfare impacts of rising

quinoa prices in Peru. World Dev. 112, 163–179. https://doi.org/10.1016/j.worlddev.2018.07.012

Bellemare, M.F., Wichman, C.J., 2020. Elasticities and the Inverse Hyperbolic Sine Transformation. Oxf. Bull. Econ. Stat. 82, 50–61. https://doi.org/10.1111/obes.12325

Beuchelt, T.D., Zeller, M., 2011. Profits and poverty: Certification's troubled link for Nicaragua's organic and fairtrade coffee producers. Ecol. Econ. 70, 1316–1324. https://doi.org/10.1016/j.ecolecon.2011.01.005

Bollen, K.A., Brand, J.E., 2010. A general panel data model with random and fixed effects: A structural equations approach. Soc. forces 89, 34. https://doi.org/10.1353/SOF.2010.0072

Bray, J.G., Neilson, J., 2017. Reviewing the impacts of coffee certification programmes on smallholder livelihoods. Int. J. Biodivers. Sci. Ecosyst. Serv. Manag. 13, 216–232. https://doi.org/10.1080/21513732.2017.1316520

Cassman, K.G., Grassini, P., 2020. A global perspective on sustainable intensification research. Nat. Sustain. 2020 34 3, 262–268. https://doi.org/10.1038/s41893-020-0507-8

Chambolle, C., Poret, S., 2013. When fairtrade contracts for some are profitable for others. Eur. Rev. Agric. Econ. 40, 835–871. https://doi.org/10.1093/ERAE/JBS039

Chen, L., Lee, H.L., 2017. Sourcing under supplier responsibility risk: The effects of certification, audit, and contingency payment. Manage. Sci. 63, 2795–2812. https://doi.org/10.1287/mnsc.2016.2466

Chiputwa, B., Spielman, D.J., Qaim, M., 2015. Food Standards, Certification, and Poverty among Coffee Farmers in Uganda. World Dev. 66, 400–412. https://doi.org/10.1016/j.worlddev.2014.09.006

Chrysafi, A., Virkki, V., Jalava, M., Sandström, V., Piipponen, J., ..., Kummu, M., 2022. Quantifying Earth system interactions for sustainable food production via expert elicitation. Nat. Sustain. 2022 510 5, 830–842. https://doi.org/10.1038/s41893-022-00940-6

Crowder, D.W., Reganold, J.P., 2015. Financial competitiveness of organic agriculture on a global scale. Proc. Natl. Acad. Sci. U. S. A. 112, 7611–7616. https://doi.org/10.1073/PNAS.1423674112/SUPPL_FILE/PNAS.1423674112.SD01.XLS

de Janvry, A., McIntosh, C., Sadoulet, E., 2015. Fair Trade and Free Entry: Can a Disequilibrium Market Serve as a Development Tool? Rev. Econ. Stat. 97, 567–573. https://doi.org/10.1162/REST_A_00512

Deaton, A., 1985. Panel data from time series of cross-sections. J. Econom. 30, 109–126.

DeFries, R.S., Fanzo, J., Mondal, P., Remans, R., Wood, S.A., 2017. Is voluntary certification of tropical agricultural commodities achieving sustainability goals for small-scale producers? A review of the evidence. Environ. Res. Lett. 12, 33001. https://doi.org/10.1088/1748-9326/aa625e

Elbers, W., van Rijsbergen, B., Bagamba, F., Hoebink, P., 2015. The impact of Utz certification on smallholder farmers in Uganda, in: Coffee Certification in East Africa: Impact on Farms, Families and Cooperatives. Wageningen Academic Publishers, Wageningen, pp. 53–81. https://doi.org/10.3920/978-90-8686-805-6_2

Ellis, R., Shockley, J., Shepherd, J., 2021. Estimating the Economic Depreciation of Farm Assets. Kentucky.

Estrella, A., Navichoc, D., Kilian, B., Dietz, T., 2022. Impact pathways of voluntary sustainability standards on smallholder coffee producers in Honduras: Price premiums, farm productivity, production costs, access to credit. World Dev. Perspect. 27. https://doi.org/https://doi.org/10.1016/j.wdp.2022.100435

Fakhfakh, I., Jarboui, A., 2020. Audit certification, earnings management and risk governance: a moderated-mediation analysis. J. Financ. Report. Account. 18, 277–299. https://doi.org/10.1108/JFRA-01-2019-0015

Fanning, A.L., O'Neill, D.W., Hickel, J., Roux, N., 2021. The social shortfall and ecological overshoot of nations. Nat. Sustain. 2021 51 5, 26–36. https://doi.org/10.1038/s41893-021-00799-z

FAO, 2023. FAOSTAT [WWW Document]. URL http://www.fao.org/faostat/en/#data/QC (accessed 12.21.20).

FAO, 2011. Guidelines for the preparation of livestock sector reviews. Animal Production and Health Guidelines. Rome, Italy.

FAO, IFAD, UNICEF, WFP, WHO, 2020. The State of Food Security and Nutrition in the World 2020. Transforming food systems for affordable healthy diets. Rome. https://doi.org/10.4060/ca9692en

Haggar, J., Soto, G., Casanoves, F., Virginio, E. de M., 2017. Environmental-economic benefits and trade-offs on sustainably certified coffee farms. Ecol. Indic. 79, 330–337. https://doi.org/10.1016/J.ECOLIND.2017.04.023

Handschuch, C., Wollni, M., Villalobos, P., 2013. Adoption of food safety and quality standards among Chilean raspberry producers – Do smallholders benefit? Food Policy 40, 64–73. https://doi.org/10.1016/J.FOODPOL.2013.02.002

Holm, S., 1979. A Simple Sequentially Rejective Multiple Test Procedure. Scand. J. Stat. 6, 65–70.

Holzapfel, S., Wollni, M., 2014. Is GlobalGAP Certification of Small-Scale Farmers Sustainable? Evidence from Thailand. J. Dev. Stud. 50, 731–747. https://doi.org/10.1080/00220388.2013.874558

Hoyle, R.H., 2023. Handbook of Structural Equation Modeling, 1st ed. Guilford.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct. Equ. Model. A Multidiscip. J. 6, 1–55. https://doi.org/10.1080/10705519909540118

Ibanez, M., Blackman, A., 2016. Is Eco-Certification a Win–Win for Developing Country Agriculture? Organic Coffee Certification in Colombia. World Dev. 82, 14–27. https://doi.org/10.1016/J.WORLDDEV.2016.01.004

INEI, DNCE, 2023. Base de Microdatos [WWW Document]. URL http://iinei.inei.gob.pe/microdatos/ (accessed 5.21.21).

International Livestock Research Institute, 2011. Gender, livestock and livelihood indicators. Nairobi.

ITC, 2023. Sustainability Map [WWW Document]. URL https://sustainabilitymap.org/trends (accessed 12.14.20).

Jena, P.R., Chichaibelu, B.B., Stellmacher, T., Grote, U., 2012. The impact of coffee certification on small-scale producers' livelihoods: A case study from the Jimma Zone, Ethiopia. Agric. Econ. 43, 429–440. https://doi.org/10.1111/j.1574-0862.2012.00594.x

Lewbel, A., 2012. Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. J. Bus. Econ. Stat. 30, 67–80. https://doi.org/10.1080/07350015.2012.643126

Marx, A., Depoorter, C., Vanhaecht, R., 2022. Voluntary Sustainability Standards: State of the Art and Future Research. Standards 2, 14–31. https://doi.org/10.3390/STANDARDS2010002

Meemken, E.M., 2021. Large farms, large benefits? Sustainability certification among family farms and agro-industrial producers in Peru. World Dev. 145, 105520. https://doi.org/10.1016/j.worlddev.2021.105520

Meemken, E.M., 2020. Do smallholder farmers benefit from sustainability standards? A systematic review and meta-analysis. Glob. Food Sec. 26. https://doi.org/10.1016/j.gfs.2020.100373

Mitiku, F., de Mey, Y., Nyssen, J., Maertens, M., 2017. Do Private Sustainability Standards Contribute to Income Growth and Poverty Alleviation? A Comparison of Different Coffee Certification Schemes in Ethiopia. Sustainability 9, 246. https://doi.org/10.3390/su9020246

Omidvar, V., Giannakas, K., 2015. The effects of fair trade on coffee growers: a framework and analysis. Agric. Econ. 46, 29–39. https://doi.org/10.1111/AGEC.12196

Otsuka, K., Fan, S., 2021. Agricultural development: New perspectives in a changing world, 1st ed. Washington, DC. https://doi.org/10.2499/9780896293830

Oya, C., Schaefer, F., Skalidou, D., 2018. The effectiveness of agricultural certification in developing countries: A systematic review. World Dev. 112, 282–312. https://doi.org/10.1016/j.worlddev.2018.08.001

Pharo, P., Oppenheim, J., Pinfield, M., Laderchi, C.R., Benson, S., Polman, P., Kalibata, A., Fan, S., Martinez, C., Samadhi, N., 2019. Growing Better: Ten Critical Transitions to Transform Food and Land Use.

Potts, J., 2007. Alternative trade initiatives and income predictability: Theory and evidence from the coffee sector . Manitoba, Canada.

Preacher, K.J., Hayes, A.F., 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav. Res. Methods 2008 403 40, 879–891. https://doi.org/10.3758/BRM.40.3.879

Qiao, Y., Halberg, N., Vaheesan, S., Scott, S., 2016. Assessing the social and economic benefits of organic and fair trade tea production for small-scale farmers in Asia: a comparative case study of China and Sri Lanka. Renew. Agric. Food Syst. 31, 246–257. https://doi.org/10.1017/S1742170515000162

Ruben, R., Fort, R., 2012. The Impact of Fair Trade Certification for Coffee Farmers in Peru. World Dev. 40, 570–582. https://doi.org/10.1016/j.worlddev.2011.07.030

Ruben, R., Fort, R., Zúñiga-Arias, G., 2009. Measuring the impact of fair trade on development. Dev. Pract. 19, 777–788. https://doi.org/10.1080/09614520903027049

Ruben, R., Zuniga, G., 2011. How standards compete: Comparative impact of coffee certification schemes in Northern Nicaragua. Supply Chain Manag. 16, 98–109. https://doi.org/10.1108/13598541111115356

Schuster, M., Maertens, M., 2013. Do private standards create exclusive supply chains? New evidence from the Peruvian asparagus export sector. Food Policy 43, 291–305. https://doi.org/10.1016/J.FOODPOL.2013.10.004

Sellare, J., 2022. New insights on the use of the Fairtrade social premium and its implications for child education. J. Rural Stud. 94, 418–428. https://doi.org/10.1016/J.JRURSTUD.2022.07.015

Sellare, J., Meemken, E.M., Kouamé, C., Qaim, M., 2020. Do Sustainability Standards Benefit Smallholder Farmers Also When Accounting For Cooperative Effects? Evidence from Côte d'Ivoire. Am. J. Agric. Econ. 102, 681–695. https://doi.org/10.1002/ajae.12015

Snider, A., Gutiérrez, I., Sibelet, N., Faure, G., 2017. Small farmer cooperatives and voluntary coffee certifications: Rewarding progressive farmers of engendering widespread change in Costa Rica? Food Policy 69, 231–242. https://doi.org/10.1016/J.FOODPOL.2017.04.009

Ssebunya, B.R., Schader, C., Baumgart, L., Landert, J., Altenbuchner, C., Schmid, E., Stolze, M., 2019. Sustainability Performance of Certified and Non-certified Smallholder Coffee Farms in Uganda. Ecol. Econ. 156, 35–47. https://doi.org/10.1016/j.ecolecon.2018.09.004

Subervie, J., Vagneron, I., 2013. A Drop of Water in the Indian Ocean? The Impact of GlobalGap Certification on Lychee Farmers in Madagascar. World Dev. 50, 57–73. https://doi.org/10.1016/J.WORLDDEV.2013.05.002

Swinnen, J., Deconinck, K., Vandemoortele, T., Vandeplas, A., 2015. Quality Standards, Value Chains, and International Development: Economic and Political Theory, 1st ed, Quality Standards, Value Chains, and International Development: Economic and Political Theory. Cambridge University

Press, Cambridge, UK. https://doi.org/10.1017/CBO9781139198912

Swinnen, J., Olper, A., Vandevelde, S., 2021. From unfair prices to unfair trading practices: Political economy, value chains and 21st century agri-food policy. Agric. Econ. 52, 771–788. https://doi.org/10.1111/AGEC.12653

Tayleur, C., Balmford, A., Buchanan, G.M., Butchart, S.H.M., Ducharme, H., Green, R.E., Milder, J.C., Sanderson, F.J., Thomas, D.H.L., Vickery, J., Phalan, B., 2017. Global Coverage of Agricultural Sustainability Standards, and Their Role in Conserving Biodiversity. Conserv. Lett. 10, 610–618. https://doi.org/10.1111/conl.12314

The World Bank, 2023. Data The World Bank [WWW Document]. URL https://data.worldbank.org/indicator/ (accessed 2.15.23).

Traldi, R., 2021. Progress and pitfalls: A systematic review of the evidence for agricultural sustainability standards. Ecol. Indic. 125, 107490. https://doi.org/10.1016/J.ECOLIND.2021.107490

UNCTAD, 2023. Voluntary Sustainability Standards in International Trade. New York, U.S.

UNFSS, 2022. Voluntary Sustainability Standards - Sustainability Agenda and Developing Countries: Opportunities and Challenges: 5th Flagship Report of the United Nations Forum on Sustainability Standards. Geneva, Switzerland.

Vagneron, I., Roquigny, S., 2011. Value distribution in conventional, organic and fair trade banana chains in the Dominican Republic. Can. J. Dev. Stud. / Rev. Can. d'études du développement 32, 324–338. https://doi.org/10.1080/02255189.2011.622619

Valkila, J., Nygren, A., 2010. Impacts of Fair Trade certification on coffee farmers, cooperatives, and laborers in Nicaragua. Agric. Human Values 27, 321–333. https://doi.org/10.1007/S10460-009-9208-7

van Rijsbergen, B., Elbers, W., Ruben, R., Njuguna, S.N., 2016. The Ambivalent Impact of Coffee

Certification on Farmers' Welfare: A Matched Panel Approach for Cooperatives in Central Kenya. World Dev. 77, 277–292. https://doi.org/10.1016/J.WORLDDEV.2015.08.021

Vanderhaegen, K., Akoyi, K.T., Dekoninck, W., Jocqué, R., Muys, B., Verbist, B., Maertens, M., 2018. Do private coffee standards 'walk the talk' in improving socio-economic and environmental sustainability? Glob. Environ. Chang. 51, 1–9. https://doi.org/10.1016/j.gloenvcha.2018.04.014

Wahba, H., Elsayed, K., 2015. The mediating effect of financial performance on the relationship between social responsibility and ownership structure. Futur. Bus. J. 1, 1–12. https://doi.org/10.1016/J.FBJ.2015.02.001

Willer, H., Sampson, G., Larrea, C., Schlatter, B., Voora, V., Bermudez, S., Dang, D., Rüger, M., Meier, C., 2022. The State of Sustainable Markets 2022: Statistics and Emerging Trends. Geneva.

Wollni, M., Andersson, C., 2014. Spatial patterns of organic agriculture adoption: Evidence from Honduras. Ecol. Econ. 97, 120–128. https://doi.org/10.1016/J.ECOLECON.2013.11.010

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data, 1st ed. The MIT Press, Cambridge, London.

Yu, J., Bouamra-Mechemache, Z., 2016. Production standards, competition and vertical relationship. Eur. Rev. Agric. Econ. 43, 79–111. https://doi.org/10.1093/ERAE/JBV004

Appendix

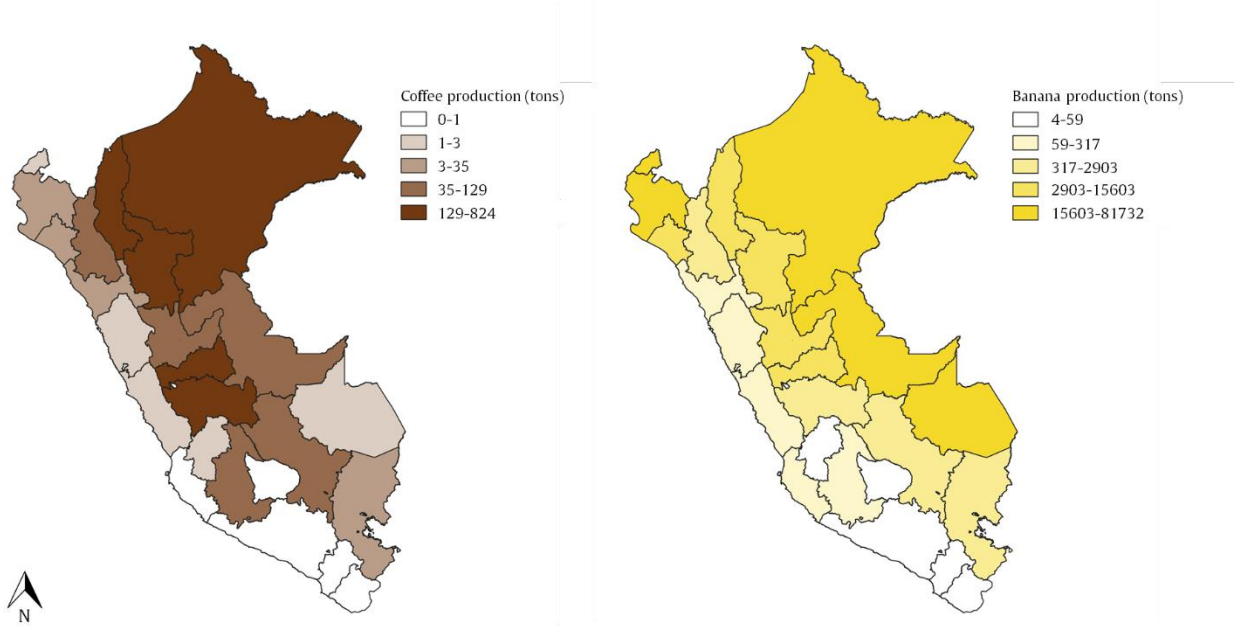


Figure A1: Coffee (left) and banana (right) production departments of Peru for 2019 based on ENA data (INEI and DNCE, 2023)

Table A1: Occurrence of a panel in the ENA dataset

	Year				
N	**2015**	**2016**	**2017**	**2018**	**2019**
380					
397					
7,268					
433					
7,398					
7,491					
558					
7,650					
7,751					
7,771					

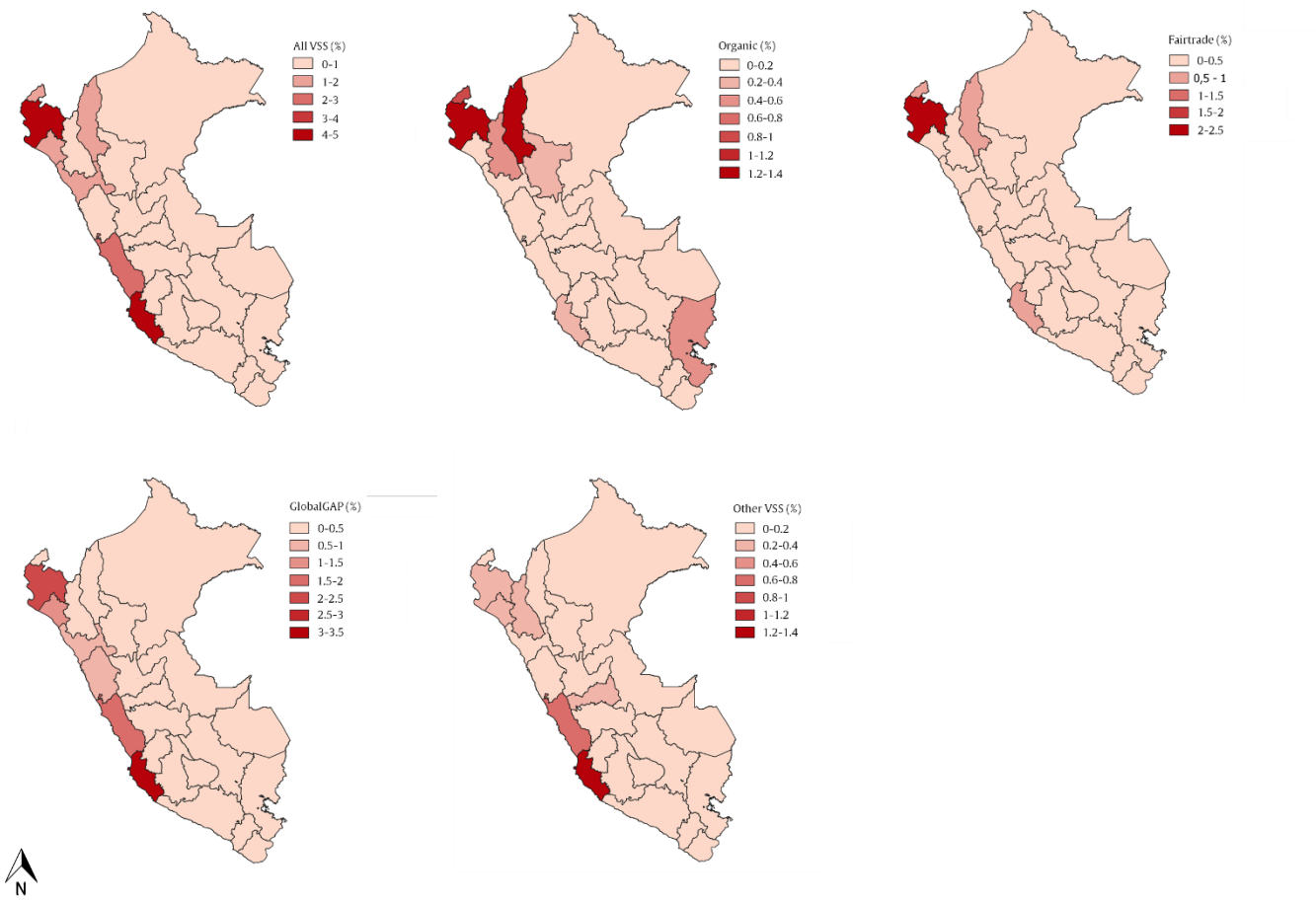


Figure A2: Importance of certification among family farms (% of sampled farms) at district-level for the period 2015-2019 based on ENA data (INEI and DNCE, 2023)

Table A2: Full unstandardized regression results of the mediation model for all crops and all VSS using the FE-FIML estimator (model 1) and including 'membership in farmer organization' as additional control variable (model 2)

Spatial level	**Farm**		**District**	
Model	**1**	**2**	**1**	**2**
Dependent variable: Net farm revenue (2019 PEN)				
Certification (β_0)	1.167 (0.465)**	1.095 (0.468)**	0.079 (1.518)	-0.124 (1.507)
Market price ($\gamma_{0,1}$)	3.163 (0.273)***	3.187 (0.272)***	2.092 (0.214)***	2.076 (0.214)***
Yield ($\gamma_{0,2}$)	0.928 (0.072)***	0.928 (0.072)***	0.640 (0.069)***	0.640 (0.069)***
Production cost ($\gamma_{0,3}$)	-1.965 (0.034)***	-1.957 (0.034)***	-1.852 (0.034)***	-1.864 (0.034)***
Total area	2.472 (0.161)***	2.491 (0.162)***	3.578 (0.107)***	3.554 (0.107)***
Total area²	-0.259 (0.036)***	-0.267 (0.036)***	-0.463 (0.022)***	-0.461 (0.022)***
Use of advanced irrigation techniques	-0.236 (0.157)	-0.246 (0.157)	0.057 (0.139)	0.059 (0.139)
Tropical livestock units	0.016 (0.003)***	0.015 (0.003)***	0.043 (0.005)***	0.042 (0.005)***
Number of household laborers on the farm	0.152 (0.037)***	0.153 (0.035)***	0.278 (0.033)***	0.292 (0.032)***
Off-farm employment by household head	0.150 (0.069)**	0.138 (0.069)**	0.167 (0.070)**	0.157 (0.070)**
Membership in farmer organization		0.133 (0.136)		0.940 (0.158)***
Dependent variable: Market price (2019 PEN/kg)				
Certification (β_1)	0.067 (0.031)**	0.067 (0.032)**	0.210 (0.073)***	0.220 (0.074)***
Total area	0.001 (0.014)	0.001 (0.014)	0.012 (0.010)	0.012 (0.010)
Total area²	-0.003 (0.003)	-0.003 (0.003)	-0.002 (0.002)	-0.002 (0.002)
Membership to farmer organization		0.000 (0.012)		0.011 (0.016)
Dependent variable: Yield (kg/ha)				
Certification (β_2)	0.043 (0.044)	0.043 (0.044)	-0.035 (0.074)	-0.025 (0.075)
Total area	-0.006 (0.004)	-0.006 (0.004)	0.004 (0.003)	0.004 (0.003)
Use of advanced irrigation techniques	0.017 (0.012)	0.016 (0.012)	0.025 (0.011)	0.025 (0.011)
Membership in farmer organization		-0.003 (0.010)		0.000 (0.012)
Dependent variable: Production cost (2019 PEN)				
Certification (β_3)	0.423 (0.115)***	0.351 (0.114)***	0.471 (0.259)*	0.441 (0.263)*
Total area	0.765 (0.026)***	0.770 (0.026)***	1.077 (0.020)***	1.067 (0.020)***
Total area²	-0.074 (0.005)***	-0.076 (0.005)***	-0.113 (0.005)***	-0.112 (0.005)***

Use of advanced irrigation techniques	0.131 (0.025)	0.126 (0.025)	0.214 (0.025)***	0.214 (0.025)***
Tropical livestock units	0.009 (0.003)***	0.009 (0.003)***	0.033 (0.005)***	0.032 (0.005)***
Membership in farmer organization		0.197 (0.024)***		0.346 (0.029)***
Mediation effects				
Market price ($\beta_1 * \gamma_{0,1}$)	0.212 (0.100)**	0.215 (0.103)**	0.439 (0.160)***	0.457 (0.161)***
Yield ($\beta_2 * \gamma_{0,2}$)	0.040 (0.041)	0.040 (0.041)	-0.022 (0.048)	-0.016 (0.048)
Production cost ($\beta_3 * \gamma_{0,3}$)	-0.831 (0.225)***	-0.687 (0.223)***	-0.873 (0.481)*	-0.822 (0.491)*
Total mediation effect ($\sum_{j=1}^{3} \beta_j * \gamma_{0,j}$)	-0.578 (0.262)**	-0.433 (0.262)*	-0.456 (0.494)	-0.194 (0.503)
N crops	49,718	49,718	44,948	44,948
Doornik-Hansen test (p-value)	0.000	0.000	0.000	0.000
F test (p-value)	0.000	0.000	0.000	0.000
Hausman test (p-value)	0.000	0.000	0.000	0.000
Cov(Certification, $\mu_{0,c,i}$)	0.007 (0.002)***	0.007 (0.002)***	0.003 (0.001)***	0.002 (0.001)***
Cov(Certification, $\mu_{1,c,i}$)	0.001 (0.000)**	0.001 (0.000)*	0.001 (0.000)***	0.001 (0.000)***
Cov(Certification, $\mu_{2,c,i}$)	0.001 (0.000)	0.001 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Cov(Certification, $\mu_{3,c,i}$)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)

*Note: *p<0.1; ** p<0.05; *** p<0.01. Clustered standard errors for the FE-FIML estimates and bootstrapped standard error for the mediation effect estimates in parentheses. We report the Doornik-Hansen test for multivariate normality, F test for unobserved heterogeneity, Hausman test for comparison with a random effects model and covariance between VSS and certain time-constant unobservable factors.*

Table A3: Unstandardized results of the single equation model estimating the total effect of VSS on net farm revenue, prices, yields and production costs

Estimator	Spatial level	Farm				District			
	Dependent variable	**Net farm revenue**	**Price**	**Yield**	**Production costs**	**Net farm revenue**	**Price**	**Yield**	**Production costs**
FE-OLS	Certification	0.332	0.056*	0.044	0.423 ***	0.134	0.286 ***	-0.029	0.487 **
		(0.481)	(0.030)	(0.043)	(0.110)	(1.481)	(0.076)	(0.071)	(0.238)
	N	49,622	21,027	49,718	49,718	44,901	25,042	44,945	44,943
	F test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Wooldridge test (p-value)	0.410	0.110	0.620	0.970	0.940	0.360	0.530	0.870
FE-FIML	Certification	0.495	0.064 **	0.041	0.431 ***	-0.960	0.225 ***	-0.026	0.505 **
		(0.516)	(0.030)	(0.043)	(0.111)	(1.533)	(0.076)	(0.073)	(0.251)
	N	49,718	49,718	49,718	49,718	44,948	44,948	44,948	44,948
	F test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Cov(Certification, $\mu_{0,c,i}$)	0.009 ***	0.001 *	0.001	-0.001	0.004 ***	0.001 ***	-0.000	0.001 ***
		(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
FD-OLS	Certification	0.530	0.056	-0.000	0.290 **	0.036	0.218 ***	-0.059	0.293
		(0.405)	(0.037)	(0.042)	(0.118)	(1.325)	(0.083)	(0.073)	(0.293)
	N	36,909	13,532	37,141	37,141	44,891	20,842	44,945	44,945
	Wooldridge test (p-value)	0.151	0.596	0.635	0.177	0.679	0.416	0.511	0.774
	Durbin-Watson test (p-value)	0.004	0.001	0.000	0.000	0.637	0.478	0.000	0.004
FD- ML	Certification	0.530	0.056	-0.000	0.290**	0.033	0.218 ***	-0.059	0.293
		(0.405)	(0.037)	(0.042)	(0.118)	(1.366)	(0.083)	(0.073)	(0.293)
	N	36,909	13,532	37,141	37,141	44,891	20,842	44,945	44,945
RE-GLS	Certification	3.438 ***	0.133 ***	0.079 *	0.519 ***	2.332 ***	0.486 ***	-0.076	0.693 ***
		(0.279)	(0.031)	(0.043)	(0.068)	(0.892)	(0.074)	(0.073)	(0.172)
	N	49,622	21,027	49,718	49,718	44,901	25,042	44,945	44,943
	Hausman test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Weight of RE estimator	0.092	0.522	0.617	0.272	0.196	0.476	0.716	0.354

RE- FIML	Certification	3.434 ***	0.129 ***	0.079 *	0.521 ***	2.845 ***	0.487 ***	-0.076	0.729 ***
		(0.279)	(0.030)	(0.043)	(0.068)	(0.866)	(0.075)	(0.073)	(0.170)
	N	49,718	49,718	49,718	49,718	44,948	44,948	44,948	44,948
	Hausman test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

*Note: *p<0.1; ** p<0.05; *** p<0.01. Clustered standard errors in parentheses. We report the F test for unobserved heterogeneity, Wooldridge test for strict exogeneity of the variable Certification, Durbin-Watson test for the estimates of the FE vs FD model, Hausman test for the estimates of the FE vs random effects model and covariance between VSS and certain time-constant unobservable factors for the FE-FIML estimates. Control variables, crop-, farm- and year- fixed effects are included but estimated coefficients not shown for brevity.*

Table A4: Unstandardized results of the single mediation models estimating effect of VSS on net farm revenue using the FE-FIML estimator

Spatial level	**Farm**			**District**		
Mediator	**Market price**	**Yield**	**Production cost**	**Market price**	**Yield**	**Production cost**
Dependent variable: Net farm revenue (2019 PEN)						
Certification (β_0)	0.206 (0.507)	0.214 (0.510)	1.248 (0.463)***	-0.795 (1.423)	-0.599 (1.412)	0.405 (1.329)
Market price ($\gamma_{0,1}$)	1.106 (0.160)***			0.813 (0.111)***		
Yield ($\gamma_{0,2}$)		0.676 (0.068)***			0.408 (0.064)***	
Production cost ($\gamma_{0,3}$)			-1.977 (0.035)***			-1.780 (0.033)***
Dependent variable: Market price (2019 PEN/kg)						
Certification (β_1)	0.059 (0.032)*			0.252 (0.086)***		
Dependent variable: Yield (kg/ha)						
Certification (β_2)		0.061 (0.045)			0.045 (0.069)	
Dependent variable: Production cost (2019 PEN)						
Certification (β_3)			0.499 (0.115)***			0.517 (0.277)*
Mediation effects						
Market price ($\beta_1 * \gamma_{0,1}$)	0.065 (0.034)*			0.205 (0.075)***		
Yield ($\beta_2 * \gamma_{0,2}$)		0.041 (0.030)			0.018 (0.028)	
Production cost ($\beta_3 * \gamma_{0,3}$)			-0.987 (0.227)***			-0.920 (0.493)*
N	43,792	43,792	43,792	44,945	44,945	44,945
Doornik-Hansen test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
F test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Hausman test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000
Cov(Certification, $\mu_{0,c,i}$)	0.009***	0.009***	0.010***	0.003***	0.004***	0.004***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Cov(Certification, $\mu_{j,c,i}$)	0.001*	-0.001	0.000***	0.001***	-0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)

*Note: *p<0.1; ** p<0.05; *** p<0.01. Clustered standard errors for the FE-FIML estimates and bootstrapped standard error for the mediation effect estimates in parentheses. Control variables, crop-, farm- and year- fixed effects are included but estimated coefficients not shown for brevity.*

Table A5: Unstandardized results of the mediation model for all crops using the FD-FIML estimator

Spatial level	**Farm**	**District**
Dependent variable: Net farm revenue (2019 PEN)		
Certification (β_0)	0.647*	0.726
	(0.384)	(1.672)
Market price ($\gamma_{0,1}$)	1.070***	0.731***
	(0.121)	(0.101)
Yield ($\gamma_{0,2}$)	1.249***	0.720***
	(0.110)	(0.112)
Production cost ($\gamma_{0,3}$)	-1.503***	-2.183***
	(0.060)	(0.063)
Dependent variable: Market price (2019 PEN/kg)		
Certification (β_1)	0.048**	0.195**
	(0.038)	(0.094)
Dependent variable: Yield (kg/ha)		
Certification (β_2)	-0.028	-0.038
	(0.045)	(0.084)
Dependent variable: Production cost (2019 PEN)		
Certification (β_3)	0.399***	0.188
	(0.123)	(0.341)
Mediation effects		
Market price ($\beta_1 * \gamma_{0,1}$)	0.052	0.143**
	(0.042)	(0.070)
Yield ($\beta_2 * \gamma_{0,2}$)	0.035	-0.027
	(0.057)	(0.061)
Production cost ($\beta_3 * \gamma_{0,3}$)	-0.600***	-0.409
	(0.185)	(0.745)
Total mediation effect ($\sum_{j=1}^{3} \beta_j * \gamma_{0,j}$)	-0.513***	-0.294
	(0.196)	(0.762)
N crops	13,532	18,631
Doornik-Hansen test (p-value)	0.000	0.000
F test (p-value)	0.000	0.000
Hausman test for FE vs RE (p-value)	0.000	0.000

*Note: *p<0.1; ** p<0.05; *** p<0.01. Clustered standard errors for the FE-FIML estimates and bootstrapped standard error for the mediation effect estimates in parentheses. We report the Doornik-Hansen test for multivariate normality, F test for unobserved heterogeneity and Hausman test for comparison with a random effects model. Control variables, crop-, farm- and year-fixed effects are included but estimated coefficients not shown for brevity.*

Table A6: Unstandardized results of the single mediation models estimating effect of VSS on the mediators and net farm revenue using the Lewbel IV-estimator

Spatial level	**Farm**				**District**			
Outcome variable	**Market price** (2019 PEN/kg)	**Yield** (kg/ha)	**Production cost** (2019 PEN)	**Net farm income** (2019 PEN)	**Market price** (2019 PEN/kg)	**Yield** (kg/ha)	**Production cost** (2019 PEN)	**Net farm income** (2019 PEN)
Certification	0.239***	0.011	1.143***	0.673	0.227**	-0.100	0.928**	2.215
	(0.109)	(0.131)	(0.425)	(0.837)	(0.102)	(0.120)	(0.361)	(1.739)
N	36,918	93,062	93,339	92,731	50,161	97,673	97,673	97,579
F test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Breusch-Pagan test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pagan-Hall test (p-value)	0.592	0.136	0.132	0.000	0.874	0.080	0.000	0.000
Hansen's J statistic (p-value)	0.204	0.725	0.616	0.257	0.484	0.113	0.252	0.327
Kleibergen-Paap rk Wald F statistic	1.315	3.143	6.023	2.775	6.962	13.697	20.561	12.917
Kleibergen-Paap rk LM statistic	5.085	17.669	20.292	14.178	12.922	12.941	11.834	15.267
Difference of two Hansen's J statistics (p-value)	0.223	0.725	0.822	0.950	0.974	0.620	0.975	0.935

*Note: *p<0.1; ** p<0.05; *** p<0.01. Clustered standard errors in parentheses. We report the F test for unobserved heterogeneity, Breusch-Pagan test for heteroscedasticity of the first stage residuals, Pagan-Hall and Hansen's J test for instrument validity, Kleibergen-Paap rk Wald statistic for weak instruments, Kleibergen-Paap rk LM statistic for under-identification, the difference of two Hansen's J statistics for endogeneity. Control variables, crop-, farm- and year-fixed effects are included but estimated coefficients not shown for brevity. Included instruments are (a subset of) the control variables: total area, tropical livestock units, the share of irrigated fields, off-farm employment of the household head, the number of household laborers, self-reported distance to the district capital, household size, age, gender, language, education, ethnicity and farm experience of the household head.*